Final Progress Report
ADAPTIVE HIGHER-ORDER METHODS FOR PROBLEMS IN ELASTODYNAMICS
Project Number DAAH04-96-1-0062

1. Foreword

The broad mission of this project was to explore new classes of numerical methods that would provide new approaches for modeling complex problems in elastostatics, elastodynamics, and wave and impact problems. We believe that two major results were obtained in the course of this work: The development of new classes of so called Discontinuous Galerkin Methods (DGMs), including the underlying mathematical theory, and Generalized Finite Element Methods (GFEMs, also referred to as hp-clouds, Partition-of-Unity Methods (PUMs)) for linear and non-linear boundary value problems. These new approaches generalize and extend so-called mesh-free techniques and are compatible with existing finite element software. At the same time, they demonstrate significant improvements in performance over traditional finite element approaches. Summaries of work done under each of these topics is given in the body of this final report.

2. Table of Contents

3. Appendix

A Posteriori Error Estimates for the Discontinuous Galerkin Method

4. Statement of the Problem Studied

The general class of problems studied in this effort were boundary-and initial-value problems in elastodynamics, but applications to broader classes were also considered. These included applications in convection-diffusion phenomena, gas dynamics, and viscous incompressible flow as modeled by the Navier-Stokes equations. The major goal of the work was to develop, analyze, and implement completely new types of methods which could overcome numerous shortcomings of existing techniques for treating large scale simulations in solid (and fluid) mechanics. Two avenues of research were pursued, one involving an attempt to extend our earlier work on Discontinuous Galerkin Methods for hyperbolic systems to problems with significant physical diffusion terms, and the second to explore new methods that extend and make competitive ideas underlying the so-called mesh-free methodologies. We believe that efforts in both of these areas were extremely successful. Further details on results and accomplishments are given in the following paragraphs.

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE
January 29, 2001

3. REPORT TYPE AND DATES COVERED
Final ~~Progress~~ Report (04/15/96-12/14/00)

4. TITLE AND SUBTITLE

ADAPTIVE HIGHER-ORDER METHODS FOR PROBLEMS IN ELASTODYNAMICS

5. FUNDING NUMBERS

DAAH04-96-1-0062

6. AUTHOR(S)

J. Tinsley Oden

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Texas Institute for Computational & Applied Mathematics
ACES Bldg. 6.412
University of Texas at Austin
Austin, Texas 78712

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

RECEIVED
APR 11 2001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 34882.12-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This document summarizes results obtained on a project aimed at developing new classes of numerical methods for the analysis of problems in elastodynamics and elastostatics. Two significant classes of new methods were developed, analyzed, and implemented:
1) the so-called hp-Cloud Method, a variant of the meshfree methods built on partitions of unity generated by traditional finite elements (also referred to as the Generalized Finite Element Method [GFEM]) and,
2) Discontinuous Galerkin Methods for broad classes of transport problems, including problems with significant diffusion. These new methods offer numerous advantages over traditional schemes for a significant class of applications. A summary of major features is given together with an Appendix outlining a priori error estimates and convergence proofs for various Discontinuous Galerkin Methods.

14. SUBJECT TERMS

15. NUMBER OF PAGES
56

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

5. Summary of Most Important Results

The success of such methods as finite elements on a broad class of engineering and scientific applications has, in many ways, made obscure several serious limitations of these approaches in treating very complex problems of interest in applications involving highly nonlinear phenomena and in problems which require high fidelity solutions. Of these limitations, first and foremost, is the a long standing problem of developing computational methods for transport phenomena: the development of cell-wise conservative schemes that can deliver very high-order local accuracy. The search for such methods has been "the holy grail" of computational fluid dynamics for over two decades and has lead to the large volume of papers concerned with very specialized techniques for hyperbolic systems that employ various non-local methods to get high-order accuracy. Until recently, none of the existing techniques had sufficient robustness or generality to be applicable to important classes of three-dimensional situations. The new DGMs not only produce local cell-wise conservative approximative schemes but they are also capable of delivering high-order polynomial accuracy. We have applied the method successfully to broad classes of simulations including convection diffusion problems, three dimensional Euler-equations for compressible flow, 2D incompressible Navier-Stokes equations, and we are in the process of implementing the algorithm for large scale problems in impact dynamics and penetration mechanics.

That the new family of Discontinuous Galerkin Methods is applicable to problems with diffusion terms was announced in a paper by Oden, Bauman, and Babuska [1] published in *the Journal of Computational Physics*. This particular method leads to an unsymmetrical system of equations for the diffusion problem which, experimentally, has proved to be robust and has the ability to support high-order local approximations. Indeed, hp versions of this technique have been developed which, with adaptivity, can yield exponential rates of convergence. The method was the basis of a dissertation of Carlos Bauman and his version of the method has become known in the literature as the Bauman-Oden Method (BOM). As noted above, the BOM has subsequently been extended to a number of significant two and three-dimensional applications, not only in computational fluid mechanics, but more recently in treating the hyperbolic coupled systems encountered in the imicible displacement equations for porous medium. Extensions of the BOM to these problems was the subject of the dissertation of Beatrice Riviera [2] and a mathematical analysis of some of its properties have been treated in a recent publication by Riviera, Gerault, and Wheeler [3]. In a recent paper, summarized in the appendix, proofs of convergence and a priori estimates for the BOM and other DGMs have been derived. We also review in the appendix to this report details of the formulation and a comparison of this method with other interior penalty methods and discontinuous methods that have been subsequently proposed. Special features of the method are listed as follows:

1. Local element polynomial approximations of arbitrary order are constructed over each cell providing a basis for local spectral type approximations.
2. Continuity conditions on the solution values and the fluxes at cell boundaries are enforced in a weak form.
3. A posteriori error indicators have been derived and used in an adaptive procedure that has allowed the implementation of the method on non-uniform hp meshes; a number test problems have been treated which demonstrate exponential convergence.
4. Most importantly, the scheme is cell-wise conservative, meaning that fluxes are balanced on each element; as far as is known by the investigators, this represents the first high-order (greater than first or second order) conservative FEM scheme.

5. A priori and a posteriori error estimates have been derived for linear applications including second order elliptic problems, convection diffusion problems in two dimensions, and first order hyperbolic problems in two dimensions. The estimated convergence rates have been confirmed in several numerical experiments. Tests indicate for p-versions of the method, sub-optimal rates are achieved. Since publication of our work, a number of new variations of the method have been proposed by other authors which include the use of penalized or stabilizing terms which may lead to robustness of the scheme and improve rates of convergence, but may also destroy the conservation properties of the method.
6. Three dimensional versions of the BOM have been developed for applications in computational fluid dynamics, (as noted above, in particular, Euler-equations and, in two dimensions, the incompressible Navier-Stokes equations). These have been applied to a significant number of benchmark problems.

We believe these methods may have important implications for coupled problems in nonlinear dynamics. Work remains to be done in deriving sharp a posteriori error estimates of these nonlinear problems, including local estimates for quantities of interest.

Mesh-free, Partition of Unity, and Generalized Finite Element Methods

Another area explored during the course of this project was the development of new families of schemes for treating both linear and nonlinear boundary value problems in mechanics and physics. Our work began on this subject in the mid-1990's with the development of so-called hp Clouds. These were techniques based on the use of Moving-Least-Squares, (MLS) techniques for deriving basis these functions for finite dimensional spaces, functions are built on overlapping domains which supported polynomials of arbitrary degree. In our earlier work, we were able to develop h, p, and hp versions of these techniques and to develop adaptive versions of these schemes that demonstrated exponential convergence.

Subsequent work showed that while these schemes could exhibit extraordinarily high rates of convergence they were not competitive with existing finite element and finite volume techniques because of the significant expense required in quadrature of the basis functions defined on spherical domains. This proved to be expensive and cumbersome and required a computational effort that dominated the entire cost of their implementation. In general, we have concluded that the cost of generating a partition of unity method on the irregular domains using the methods of least square, a common step in most of the mesh-free methods now in use, is prohibitive.

In 1997 and 1998, we began exploring the use of conventional finite element methods as a device to generate a partition of unity over a given computational domain. Such an approach automatically overcame the quadrature problem, since quadrature points were already embedded in the finite element master elements, but they (FEMs) provided nodal basis functions on overlapping domains on which high-order polynomial approximations and other types of Treffz approximations could be constructed. While the resulting methods could not qualify as "mesh-free," they possess a number of very desirable properties which ultimately proved to be superior to conventional finite element techniques. We have termed these methods GFEMs: Generalized Finite Element Methods. They were first reported in papers by Oden, Duarte, and Zienkiewicz, and in a one-dimensional case, by Babuska and Melenik [4]. With the addition of special functions, these techniques can provide accelerated rates of

convergence in cases where traditional finite element methods diverge or converge very slowly. These include problems in which one has very rough highly oscillatory coefficients, boundary layers, interfaces, singularities, and other local features that influence convergence.

One of the most important applications of the GFEM has been in the treatment of crack problems and propagation of cracks through domains. We have developed techniques for allowing cracks to occur in a given domain and to propagate through the domain without altering the original mesh. Commercial engineering software companies have expressed interest in our results and one (Altair Engineering) has used these methodologies in applications involving two and three-dimensional crack propagation problems. Apparently, our techniques have formed the basis of certain versions of commercial products that are or soon will be on the market.

In many of the applications and experimental tests of these GFEMs, performances have been observed which are considerably faster and more efficient than traditional finite element methods. As noted earlier, one can also cite problems in which these methods exhibit extraordinarily high convergence rates while FEMs do not converge at all. The subject of GFEM has been further explored in by dissertation by Kevin Copps [5]; it has been the subject of a NSF workshop and is a popular topic in conferences on computational science and engineering. The full exploitation of these methods in treating significant non-linear problems remains to be explored. We feel that one of the most important features of GFEMs over some of the recently reported mesh-free methods is that they can be easily embedded into existing finite element codes. The GFEM, in other words, can be used to add to existing finite element packages the ability to treat a long list of special features, particularly the propagation of cracks across fixed meshes and the treatment of singularities in boundary layer effects. We have derived a priori and a posteriori error estimates for these methods and have developed adaptive versions of them.

One area relative to GFEMs that remains to be more thoroughly explored is the development of preconditioners for the method for very large problems, including the use of the techniques in a parallel computer environment. Since the GFEM technique involves superimposing polynomial type approximations (or, in some cases, special functions) on a model-based partition of unity generated by traditional FEMs, in most cases rank deficient stiffness matrices are produced. Thus, special techniques must be developed to solve the resulting linear systems. We have developed and implemented one technique that seems to work satisfactorily, but a great deal of additional theoretical work needs to be done to develop fully robust iterative techniques for handling the linear systems generated by GFEMs. We believe this problem is solvable, and certainly is an area worth further study.

6. List of all Publications and Technical Reports

<u>Number of Papers published in Journals and Conference proceedings:</u>
12/96-12/00 = 21

<u>Refereed Journal Publications</u>:

1. Duarte, C.A.M. and Oden, J.T. "An hp Adaptive Method Using Clouds." Computer Methods in Applied Mechanics and Engineering, Special Issue on Meshless Methods. Edited by T. Belytschko, W.K. Liu, and J.T. Oden, volume 139, pp. 237-262, 1996.

2. Duarte, C. Armando and Oden, J.T. "Hp Clouds - A Meshless Method to Solve Boundary-Value Problems." Numerical Methods for Partial Differential Equations, volume 12, pp. 1-33, 1996.
3. Duarte, C. Armando and Oden, J. Tinsley, "An h-p Adaptive Method Using Clouds." Computer Methods in Applied Mechanics and Engineering, volume 139, 237-262, 1996.
4. Oden, J.T.; Duarte, C.A.M.; and Zienkiewicz, O.C. "A New Cloud-Based hp Finite Element Method", Computational Methods and Applied Mechanical Engineering, volume 153, number 1-2, pp. 117-126, 1998.
5. Oden, J. Tinsley; Feng, Yusheng; and Prudhomme, Serge. "Local and Pollution Error Estimation for Stokesian Flows." International Journal for Numerical Methods in Fluids, volume 27, pp. 33-39, 1998.
6. Messina, Paul; Culler, David; Pfieffer, Wayne; Martin, W.; Oden, J. Tinsley; and Smith, Gary. "Architecture." Communications of the ACM, volume 41, number 11, pp. 26-44, November 1998.
7. Babuska, Ivo; Baumann, Carlos Erik; and Oden, J. Tinsley. "A Discontinuous hp Finite Element Method for Diffusion Problems: 1-D Analysis", Computers and Mathematics with Applications, volume 37, pp. 103-122, 1999.
8. Baumann, Carlos and Oden, J. Tinsley. "A Discontinuous hp Finite Element Method for Convection-Diffusion Problems." Special Issue: <u>Spectral, Spectral Element, and HP Methods in CFD</u>, edited by George Karniadakis. Computer Methods in Applied Mechanics and Engineering, volume 175, pp. 311-341, 1999.
9. Baumann, Carlos Erik and Oden, J. Tinsley. "A Discontinuous hp Finite Element Method for the Solution of the Euler and Navier-Stokes Equations." Special Issue in memory of Dick Gallagher, edited by Juan Heinrich. International Journal for Numerical Methods in Fluids, volume 31, pp. 79-95, 1999.
10. Oden, J. Tinsley; Babuska, Ivo; and Baumann, Carlos Erik. "A Discontinuous hp Finite Element Method for Diffusion Problems", Journal of Computational Physics, volume 146, pp. 491-519, 1998.
11. Baumann, Carlos Erik and Oden, J. Tinsley. "An Adaptive-Order Discontinuous Galerkin Method for the Solution of the Euler Equations of Gas Dynamics." Special Issue in memory of Dick Gallagher, edited by Roland Lewis. International Journal for Numerical Methods in Engineering, volume 47, pp. 61-73, 2000.
12. Duarte, C.A.; Babuska, I.; and Oden, J.T. "Generalized Finite Element Methods for Three Dimensional Structural Mechanics Problems." Computers and Structures (to appear)
13. Korneev, V.; Flaherty, J.E.; Oden, J. T.; and Fish, J. "Additive Schwarz Algorithms for solving hp-Version Finite Element Systems on Triangular Meshes." SIAM Journal on Numerical Analysis (in review)
14. Prudhomme, Serge; Pascal, Frederic; Oden, J. Tinsley; and Romkes, Albert. "A priori error estimate for the Baumann-Oden version of the discontinuous Galerkin method", in C.R. Acad.Sci. Paris, I, Numerical Analysis (submitted December 2000)

<u>**Conference Proceedings:**</u>

1. Duarte, C. Armando and Oden, J.T. "A New Meshless Method to Solve Boundary-Value Problems." Proceedings of the XVI CILAMCE-Iberian Latin American Conference on Computational Methods for Engineering, held in Curitiba, Brazil, November 1995. Edited by R.D. Machado, pp. 90-99, 1996.
2. Oden, J.T. and Duarte, C.A. "Clouds, Cracks, and FEM's." Recent Developments in Computational and Applied Mechanics, B.D. Reddy (Ed.), CIMNE, Barcelona, pp. 302-321, 1997.
3. Baumann, Carlos Erik and Oden, J. Tinsley, "A Discontinuous hp Finite Element Method for the Solution of the Euler Equations of Gas Dynamics." Proceedings of

the Tenth International Conference on Finite Elements in Fluids. Edited by M. Hafez and J.C. Heinrich. Held January 5-8, 1998 in Tucson AZ,. pp. 437-443, 1998.

4. Baumann, Carlos Erik and Oden, J. Tinsley, "A Discontinuous hp Finite Element Method for the Solution of the Navier-Stokes Equations", Proceedings of the Tenth International Conference on Finite Elements in Fluids. Edited by M. Hafez and J.C. Heinrich. Held January 5-8, 1998 in Tucson AZ, pp. 162-68, 1998
5. Duarte, C.A.; Babuska, I.; and Oden, J.T. "Generalized Finite Element Methods for Three-Dimensional Structural Mechanics Problems." Modeling and Simulation Based Engineering, edited by S.N. Atluri and P.E. O'Donoghue, volume 1, pp. 53-58, 1998. Proceedings of the International Conference on Computational Engineering Science held October 5-9, 1998 in Atlanta, GA.
6. Oden, J.T. and Baumann, Carlos. "A Conservative DGM for Convection-Diffusion and Navier-Stokes Problems", Proceedings of the DGM International Symposium, held May 24-26, 1999 in Providence, RI.

Technical Reports:

1. Oden, J.T.; Duarte, C.A.; and Zienkiewicz, O.C. "A New Cloud-Based hp-Finite Element Method", TICAM Report 96-55, December 1996.
2. Oden, J. Tinsley; Babuska, Ivo; and Baumann, Carlos Erik, "A Discontinuous hp Finite Element Method for Diffusion Problems." TICAM Report 97-21, November 1997.
3. Babuska, Ivo; Baumann, Carlos Erik; and Oden, J. Tinsley, "A Discontinuous hp Finite Element Method for Diffusion Problems: 1-D Analysis", TICAM Report 97-22, November 1997.

Other Publications:

Book Chapters:

1. Oden, J.T. and Baumann, Carlos. "A Conservative DGM for Convection-Diffusion and Navier-Stokes Problems", Discontinuous Galerkin Methods: Theory, Compilation and Applications. Lecture Notes in Computational Science and Engineering, edited by B. Cockburn, G.E. Karniadakis, C.W. Shu. v. 11, pp. 179-196, 1999.
2. Oden, J.T. "Discontinuous Galerkin Methods in the Solution of the Convection Diffusion Equation." Appendix B, volume 3, of Finite Elements in Engineering, by O.C. Zienkiewicz and R. Taylor, Fifth Edition, Butterworths Publishing, London 2000.

Books (Edited, Co-Edited):

1. Belytschko, T.; Liu, W.K.; and Oden, J.T. Meshless Methods: Special Issue, Computer Methods in Applied Mechanics and Engineering, Elsevier Science Publishers, Amsterdam, 1996.
2. Ainsworth, M. and Oden, J.T. A Posteriori Error Estimation in Finite Element Analysis, Wiley-Interscience, N.Y., 2000.

7. Students Supported

Bauman, Carlos	Ph.D completed 1998
Duarte, Armando	Ph.D completed 1996
Prudhomme, Serge	Ph.D completed May, 1999
Romkes, Albert	Ph.D in progress
Vemaganti, Kumar	Ph.D completed Dec., 2000

APPENDIX

1. Introduction

There has been renewed interest in Discontinuous Galerkin Methods (DGM) recently, primarily due to the discovery that variants of these methods could be used effectively to solve diffusion problems as well as problems of pure convection. One such DGM was presented in the dissertation of Baumann [7] and reported in the paper of Oden, Babuška, and Baumann [18]; summary of other versions of DGMs and a lengthy historical review of this subject can be found in the record volume edited by Cockburn, Karniadakis, and Shu [10]. The DGM possesses a number of important properties that set them apart from traditional conforming Galerkin-finite element methods: they are elementwise conservative, can support high order local approximations that can vary nonuniformly over the mesh, are readily parallelizable, and, for time-dependent problems, lead to block-diagonal mass matrices, even for high-order polynomial approximations. These properties make DGMs attractive candidates for a broad collection of applications.

Several papers have been published in the mathematical literature on *a priori* error estimates for various DGMs for diffusion problems. In particular, an analysis of one-dimensional versions of the Baumann-Oden method was reported by Babuška, Oden, and Baumann [2]. Error estimates for several types of DGMs and for the related Internal Penalty Galerkin Methods were presented in the dissertation of Rivière [19] and in the paper of Rivière, Wheeler, and Girault [20]. Several other studies on *a priori* error estimates for DGMs have appeared recently; see, for example, the report of Chen [9] and the analysis of Süli, Schwab, and Houston [22,15]. Convergence analysis of other variants of DGM can be found in [10].

Here we present a detailed derivation of *a priori* error estimates for several *hp*-versions of DG-finite element methods for linear diffusion problems (the Poisson problem) on two-dimensional domains. In some cases, important steps in our analysis follows the approach of Rivière, Wheeler, and Girault [20], but other steps differ in detail. We present a series of approaches in which different versions of DGMs, including those with penalty terms, can be analyzed. Our final estimates differ in predicted rates of convergence with respect to the polynomial degree p obtained in [20,19] and reflect rates consistent with the computed results of Baumann [7].

2. Notations and Preliminaries

We shall choose the domain Ω as a bounded open set in $\mathbb{R}^2$, with Lipschitz continuous boundary $\partial\Omega$. We will denote Γ_D the part of the boundary $\partial\Omega$ on which Dirichlet conditions are prescribed and Γ_N the part on which Neumann conditions are prescribed. Formally, the boundary $\partial\Omega$ is decomposed into the parts Γ_D and Γ_N such that $\bar{\Gamma}_D \cup \bar{\Gamma}_N = \partial\Omega$, and $\Gamma_D \cap \Gamma_N = \varnothing$.

2.1. Finite Element Partition

Let $\mathcal{P}_h$ denote a partition of the domain Ω, i.e. $\mathcal{P}_h$ is a finite collection of N_e open subdomains (elements) K_i, $i = 1, 2, \dots, N_e$, such that:

$$\overline{\Omega} = \bigcup_{K_i \in \mathcal{P}_h} \overline{K_i}, \qquad \text{and} \quad K_i \cap K_j = \varnothing, \quad i \neq j.$$

The size and shape of an element K_i, or simply K, of $\mathcal{P}_h$ are measured in terms of two quantities, h_K and ρ_K, defined as:

$$h_K = \text{diam}(K),$$
$$\rho_K = \sup\{\text{diam}(\mathcal{B}); \mathcal{B} \textit{ is a ball contained in } K\}.$$

We also introduce the parameter h associated with the partition $\mathcal{P}_h$:

$$h = \max_{K \in \mathcal{P}_h} h_K. \tag{2.1}$$

Definition *A family $\{\mathcal{P}_h\}$ of partitions $\mathcal{P}_h$ is said to be shape regular as h tends to zero if there exists a number $\varrho > 0$, independent of h and K such that:*

$$\frac{h_K}{\rho_K} \leq \varrho, \qquad \forall K \in \mathcal{P}_h. \tag{2.2}$$

In this appendix, all partitions $\mathcal{P}_h$ are assumed to be shape-regular.

In addition, we shall associate with each element K the element boundary ∂K. The unit normal vector outward from K (resp. K_i) is denoted by $\mathbf{n}$ (resp. $\mathbf{n}|_i$).

Given a partition $\mathcal{P}_h$, we shall denote the collection of edges of $\mathcal{P}_h$ (points in one dimension, faces in three dimensions) by the set $\mathcal{E}_h = \{\gamma_l\}$, $l = 1, \dots, N_\gamma$. Edges represent here open subsets of either Ω or $\partial\Omega$. We thus introduce the set Γ_{int} of interior edges as:

$$\Gamma_{int} = \left(\bigcup_{l=1}^{N_\gamma} \gamma_l\right) \setminus \partial\Omega \tag{2.3}$$

so that:

$$\bigcup_{l=1}^{N_\gamma} \bar{\gamma}_l = \bar{\Gamma}_D \cup \bar{\Gamma}_N \cup \bar{\Gamma}_{int}.$$

In the same way, we shall decompose $\mathcal{E}_h$ into three subsets as:

$$\mathcal{E}_h = \mathcal{E}_{h,D} \cup \mathcal{E}_{h,N} \cup \mathcal{E}_{h,int}.$$

Then, $\gamma \in \mathcal{E}_{h,D}$ if it lies on Γ_D, and $\gamma \in \mathcal{E}_{h,N}$ if it lies on Γ_N. Moreover, as shown in Fig. 1, $\gamma_{ij} \in \mathcal{E}_{h,int}$ denotes an edge (interface) between two adjacent elements K_i and K_j, where by convention $i > j$. For each edge γ, we also associate a unit normal vector $\mathbf{n}$. In the case γ is an edge associated with an element K_i adjacent to $\partial\Omega$, i.e. $\gamma \in \mathcal{E}_{h,D} \cup \mathcal{E}_{h,N}$, the unit normal vector is simply defined as $\mathbf{n} = \mathbf{n}|_i$. For an interior edge $\gamma_{ij} \in \mathcal{E}_{h,int}$, with the convention $i > j$, $\mathbf{n}$ is chosen as the unit normal vector outward from K_i, so that $\mathbf{n} = \mathbf{n}|_i = -\mathbf{n}|_j$ (see Fig. 1). In subsequent analyses, C will denote generic positive constants, not necessarily the same in different places.

Remark 1 *Using simple geometrical properties, one can show that each edge γ in a shape-regular partition satisfies:*

$$\frac{1}{\varrho} h_K \leq \rho_K \leq |\gamma| \leq h_K, \tag{2.4}$$

where $|\gamma|$ denotes the length of γ. In other words, h_K and γ are equal within a constant. Therefore, we will interchangeably use h_K or γ (preferably h_K).

2.2. Spaces

Let s be a positive integer. For any given open set S (S may define the whole domain Ω, an element K of $\mathcal{P}_h$, or an edge γ of $\mathcal{E}_h$), the spaces $H^s(S)$ will denote the usual Sobolev spaces with norm $\|\cdot\|_{s,S}$. In the particular case in which S represents Ω, the norm will simply be denoted $\|\cdot\|_s$. Moreover, $H_0^1(S)$ is the set of functions in $H^1(S)$ which vanish on the boundary ∂S of S, i.e.

$$H_0^1(S) = \{v \in H^1(S); v = 0 \text{ on } \partial S\},$$

and $H(\text{div}, S)$ denotes the space:

$$H(div, S) = \{\mathbf{v} \in (L^2(S))^2; \nabla \cdot \mathbf{v} \in L^2(S)\}.$$

The so-called (mesh-dependent) *broken space* $H^s(\mathcal{P}_h)$ will be defined as:

$$H^s(\mathcal{P}_h) = \{v \in L^2(\Omega);\ v|_K \in H^s(K),\ \forall K \in \mathcal{P}_h\}.$$

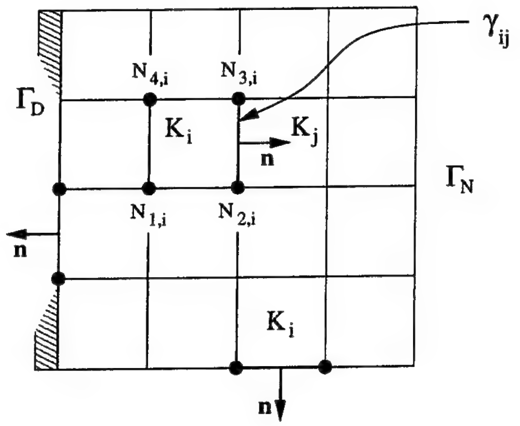

Figure 1. Element interface γ_{ij} and unit normal vector **n**.

The norm associated with the space $H^s(\mathcal{T}_h)$ is given as:

$$\|v\|_{s,\mathcal{T}_h} = \left(\sum_{K \in \mathcal{T}_h} \|v\|^2_{s,K} \right)^{1/2}$$

where $\|v\|_{s,K}$ is the Sobolev norm on K.

We will consider finite element spaces $\mathcal{V}^{hp}$ of polynomial functions, possibly discontinuous at the element interfaces, such as:

$$\mathcal{V}^{hp} = \{v \in L^2(\Omega);\ v|_K = \hat{v} \circ F_K^{-1}, \hat{v} \in P_p(\widehat{K}),\ \forall K \in \mathcal{T}_h\} \tag{2.5}$$

where F_K is the affine mapping from the master element $\widehat{K}$ to the element K in the partition, and $P_p(\widehat{K})$ is the space of polynomial functions of degree at most p on $\widehat{K}$.

In *hp* methods, the polynomial degree can actually vary from one element to the other. Denoting p_K the polynomial degree associated with the element K, we define the global value p for the partition $\mathcal{T}_h$ as:

$$p = \min_{K \in \mathcal{T}_h} p_K. \tag{2.6}$$

One advantage of DGMs over conventional *hp* finite element methods is that the polynomial degrees p_K do not necessarily match at the interfaces of the elements.

3. Formulations for the Poisson Model Problem

3.1. Model Problem

We shall consider here the following Poisson model problem: find the scalar function u which is the solution of

$$-\Delta u + cu = f, \quad \text{in } \Omega, \tag{3.1}$$

and which satisfies the boundary conditions:

$$\begin{aligned} u &= u_0, && \text{on } \Gamma_D, \\ \mathbf{n} \cdot \nabla u &= g, && \text{on } \Gamma_N. \end{aligned} \tag{3.2}$$

Here $f \in L^2(\Omega)$ represents the load scalar and c is a positive constant over the domain Ω.

We now proceed with the derivation of weak formulations of the Poisson equation on discontinuous spaces. Let u, for the moment, be a sufficiently smooth function. The regularity of u shall be discussed later in the appendix, namely in Subsection 3.3. Multiplying (3.1) by a function v in $H^2(\mathcal{P}_h)$ and integrating over the domain Ω, we obtain:

$$\int_\Omega (-\nabla \cdot \nabla u + cu)\, v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x.$$

Unlike the classical continuous finite element approach, we shall first decompose the integrals in the above equation into element contributions

$$\sum_{K \in \mathcal{P}_h} - \int_K (\nabla \cdot \nabla u)\, v \,\mathrm{d}x + \sum_{K \in \mathcal{P}_h} \int_K cuv \,\mathrm{d}x = \sum_{K \in \mathcal{P}_h} \int_K f v \,\mathrm{d}x,$$

and then integrate by parts, so that:

$$\sum_{K \in \mathcal{P}_h} \int_K (\nabla u \cdot \nabla v + cuv) \mathrm{d}x - \sum_{K \in \mathcal{P}_h} \int_{\partial K} (\mathbf{n} \cdot \nabla u)\, v \,\mathrm{d}s = \sum_{K \in \mathcal{P}_h} \int_K f v \,\mathrm{d}x. \tag{3.3}$$

We observe that the boundary integrals are defined on each element boundary; those are now splitted according to the type of boundary such as:

$$\begin{aligned} \sum_{K \in \mathcal{P}_h} \int_{\partial K} (\mathbf{n} \cdot \nabla u)\, v \,\mathrm{d}s = & \sum_{\gamma \in \mathcal{E}_{h,D}} \int_\gamma (\mathbf{n} \cdot \nabla u)\, v \,\mathrm{d}s \\ & + \sum_{\gamma \in \mathcal{E}_{h,N}} \int_\gamma (\mathbf{n} \cdot \nabla u)\, v \,\mathrm{d}s \\ & + \sum_{\gamma_{ij} \in \mathcal{E}_{h,int}} \int_{\gamma_{ij}} (\mathbf{n} \cdot \nabla u)_i\, v_i + (\mathbf{n} \cdot \nabla u)_j\, v_j \,\mathrm{d}s. \end{aligned}$$

where v_i and v_j denote the restrictions of v on the elements K_i and K_j respectively. In the same way, $(\mathbf{n}\cdot\nabla u)_i$ and $(\mathbf{n}\cdot\nabla u)_j$ represents the restrictions of the flux $\mathbf{n}\cdot\nabla u$ on K_i and K_j.

In general, except occasionally to avoid confusion, we shall simplify the notation of these boundary integrals, by rewriting them, for instance,

$$\sum_{\gamma\in\mathcal{E}_{h,D}} \int_\gamma (\mathbf{n}\cdot\nabla u)\, v\,\mathrm{d}s = \int_{\Gamma_D} (\mathbf{n}\cdot\nabla u)\, v\,\mathrm{d}s,$$

$$\sum_{\gamma\in\mathcal{E}_{h,N}} \int_\gamma (\mathbf{n}\cdot\nabla u)\, v\,\mathrm{d}s = \int_{\Gamma_N} (\mathbf{n}\cdot\nabla u)\, v\,\mathrm{d}s.$$

Moreover, the treatment of the *interior* boundary integrals is as follows. Given an edge $\gamma_{ij}\in\mathcal{E}_{h,int}$ shared by two adjacent elements K_i and K_j, $i>j$, we first note that:

$$(\mathbf{n}\cdot\nabla u)_i\, v_i + (\mathbf{n}\cdot\nabla u)_j\, v_j = \mathbf{n}\cdot(\nabla u)_i\, v_i - \mathbf{n}\cdot(\nabla u)_j\, v_j,$$

where $\mathbf{n}$ is now the unit normal vector with respect to the edge γ_{ij} as defined in the previous section. By analogy with the formula below where a,b,c and d are real numbers:

$$ac - bd = \frac{1}{2}(a+b)(c-d) + \frac{1}{2}(a-b)(c+d), \tag{3.4}$$

we can write the integrand as:

$$\begin{aligned}
&\mathbf{n}\cdot(\nabla u)_i\, v_i - \mathbf{n}\cdot(\nabla u)_j\, v_j \\
&\quad = \frac{1}{2}\left(\mathbf{n}\cdot(\nabla u)_i + \mathbf{n}\cdot(\nabla u)_j\right)(v_i - v_j) + \frac{1}{2}\left(\mathbf{n}\cdot(\nabla u)_i - \mathbf{n}\cdot(\nabla u)_j\right)(v_i + v_j) \\
&\quad = \langle \mathbf{n}\cdot\nabla u\rangle\,[v] + [\mathbf{n}\cdot\nabla u]\,\langle v\rangle\,.
\end{aligned}$$

Here $[v]$ and $\langle v\rangle$ respectively denote the jump and average of v on an interior edge γ_{ij}, $i>j$, of any function $v\in H^s(K_i)\times H^s(K_j)$, $s>1/2$, i.e.

$$[v] = v_i - v_j,$$

$$\langle v\rangle = \frac{1}{2}(v_i + v_j).$$

We conveniently extend the definition of $[v]$ and $\langle v\rangle$, following Chen [9], to an edge γ lying on Γ_D as:

$$[v] = v,$$

$$\langle v\rangle = v.$$

It allows us to combine the interior and Dirichlet boundary terms in only one integral as:

$$\sum_{\gamma_{ij}\in\mathcal{E}_{h,int}} \int_{\gamma_{ij}} (\mathbf{n}\cdot\nabla u)_i\, v_i + (\mathbf{n}\cdot\nabla u)_j\, v_j\,\mathrm{d}s + \sum_{\gamma\in\mathcal{E}_{h,D}} \int_{\gamma} (\mathbf{n}\cdot\nabla u)\, v\,\mathrm{d}s$$
$$= \int_{\Gamma_{int}\cup\Gamma_D} \langle\mathbf{n}\cdot\nabla u\rangle\,[v] + [\mathbf{n}\cdot\nabla u]\,\langle v\rangle\,\mathrm{d}s.$$

Remark 2 *Note that when $u\in H^2(\Omega)$, the fluxes $[\mathbf{n}\cdot\nabla u]$ are continuous almost everywhere in Ω, which yields*

$$\int_{\Gamma_{int}} [\mathbf{n}\cdot\nabla u]\,\langle v\rangle\,\mathrm{d}s = 0, \qquad \forall v\in H^2(\mathcal{T}_h). \tag{3.5}$$

Consequently, (3.3) can now be reduced, when $u\in H^2(\Omega)$ and applying the Neumann boundary condition, to:

$$\sum_{K\in\mathcal{T}_h} \int_K (\nabla u\cdot\nabla v + cuv)\,\mathrm{d}x - \int_{\Gamma_{int}\cup\Gamma_D} \langle\mathbf{n}\cdot\nabla u\rangle\,[v]\,\mathrm{d}s = \sum_{K\in\mathcal{T}_h}\int_K fv\,\mathrm{d}x + \int_{\Gamma_N} gv\,\mathrm{d}s.$$

We introduce the following bilinear form $B(\cdot,\cdot)$ defined on $H^2(\mathcal{T}_h)\times H^2(\mathcal{T}_h)$ and the linear form $L(\cdot)$ defined on $H^2(\mathcal{T}_h)$ such as:

$$B(u,v) = \sum_{K\in\mathcal{T}_h}\int_K (\nabla u\cdot\nabla v + cuv)\,\mathrm{d}x, \tag{3.6}$$

$$F(v) = \sum_{K\in\mathcal{T}_h}\int_K fv\,\mathrm{d}x + \int_{\Gamma_N} gv\,\mathrm{d}s. \tag{3.7}$$

We also consider the bilinear form $J(\cdot,\cdot)$ on $H^2(\mathcal{T}_h)\times H^2(\mathcal{T}_h)$, which incorporates all boundary integrals on Γ_{int} and Γ_D, as:

$$J(u,v) = \int_{\Gamma_{int}\cup\Gamma_D} \langle\mathbf{n}\cdot\nabla u\rangle\,[v]\,\mathrm{d}s. \tag{3.8}$$

Then, a general discontinuous weak formulation of the Poisson equation reads:

$$B(u,v) - J(u,v) = F(v), \qquad \forall v\in H^2(\mathcal{T}_h). \tag{3.9}$$

This above variational form constitutes the starting point to derive formulations of various Discontinuous Galerkin Finite Element Methods (concisely, DGMs.)

3.2. Weak Formulations and Finite Element Discretizations

All the formulations presented below use the observation that, for $u \in H^1(\Omega) \cap H^2(\mathcal{T}_h)$, the jump $[u]$ vanishes on each γ_{ij}:

$$\int_{\gamma_{ij}} v\,[u]\,\mathrm{d}s = 0, \qquad \forall v \in L^2(\gamma_{ij}). \tag{3.10}$$

It follows that:

$$\int_{\Gamma_{int}} \langle \mathbf{n} \cdot \nabla v \rangle\,[u]\,\mathrm{d}s = 0, \qquad \forall v \in H^2(\mathcal{T}_h). \tag{3.11}$$

Moreover, the Dirichlet boundary condition can be applied in the following weak manner:

$$\int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\,u\,\mathrm{d}s = \int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\,u_0\,\mathrm{d}s, \qquad \forall v \in H^2(\mathcal{T}_h). \tag{3.12}$$

Therefore, introducing the linear form $J_0(\cdot)$ defined as:

$$J_0(v) = \int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\,u_0\,\mathrm{d}s, \qquad \forall v \in H^2(\mathcal{T}_h), \tag{3.13}$$

we observe that, for $u \in H^1(\Omega) \cap H^2(\mathcal{T}_h)$ and $u = u_0$ on Γ_D,

$$J(v,u) = J_0(v), \qquad \forall v \in H^2(\mathcal{T}_h). \tag{3.14}$$

3.2.1. Global Element Method - GEM

Introducing the bilinear form $\mathcal{B}_-(\cdot,\cdot)$, the subscript $-$ referring to the fact that we substract the term $J(v,u)$ to the left hand side of (3.9), and the linear form $\mathcal{F}_-(\cdot)$

$$\begin{aligned} \mathcal{B}_-(u,v) &= B(u,v) - J(u,v) - J(v,u), \\ \mathcal{F}_-(v) &= F(v) - J_0(v), \end{aligned} \tag{3.15}$$

the Global Element Method consists in finding u such that:

$$\mathcal{B}_-(u,v) = \mathcal{F}_-(v), \qquad \forall v \in H^2(\mathcal{T}_h). \tag{3.16}$$

One advantage of this method is that it defines a symmetric problem. On the other hand, a significant disadvantage is that the bilinear form is not guaranteed to be semi-positive definite. When dealing with time-dependent problems, this could imply that some eigenvalues have negative real parts, causing the formulation to be unconditionally unstable.

The corresponding finite element discretization of the above problem consists in finding $u_h \in \mathcal{V}^{hp}$ such that:

$$\mathcal{B}_-(u_h,v) = \mathcal{F}_-(v), \qquad \forall v \in \mathcal{V}^{hp}. \tag{3.17}$$

This method was introduced by Delves *et al.* [11–14] with the particular objective of accelerating convergence of iterative schemes.

3.2.2. Symmetric Interior Penalty Galerkin Method - SIPG

To enforce stability of the discontinuous method, i.e. continuity of the solution at the interface of the elements, penalty terms have been added to the formulation by Arnold [1] and Wheeler [23]. Let us introduce the following penalty terms:

$$J^{\sigma}(u,v) = \sum_{\gamma_{ij}\in\mathcal{E}_{h,int}} \int_{\gamma_{ij}} \sigma\,[u]\,[v]\,\mathrm{ds} + \sum_{\gamma\in\mathcal{E}_{h,D}} \int_{\gamma} \sigma u v\,\mathrm{ds} = \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[u]\,[v]\,\mathrm{ds},$$

and

$$J_0^{\sigma}(v) = \sum_{\gamma\in\mathcal{E}_{h,D}} \int_{\gamma} \sigma u_0 v\,\mathrm{ds} = \int_{\Gamma_D} \sigma u_0 v\,\mathrm{ds},$$

where σ represents the penalty parameter which depends on the length of the edges γ_{ij} and γ and the polynomial degree used in the elements; namely $\sigma = \sigma(h,p)$. Then the SIPG method is similar to the GEM except for the penalty terms. Indeed, introducing the forms:

$$\begin{aligned}\mathcal{B}_{-}^{\sigma}(u,v) &= B(u,v) - J(u,v) - J(v,u) + J^{\sigma}(u,v),\\ \mathcal{F}_{-}^{\sigma}(v) &= F(v) - J_0(v) + J_0^{\sigma}(v),\end{aligned} \tag{3.18}$$

the Symmetric Interior Penalty Galerkin problem is to find u such that:

$$\mathcal{B}_{-}^{\sigma}(u,v) = \mathcal{F}_{-}^{\sigma}(v), \qquad \forall v \in H^2(\mathcal{P}_h). \tag{3.19}$$

Note that when σ takes on the value zero, we naturally retrieve the GE method.

The finite element analogue of problem (3.19) is to find $u_h \in \mathcal{V}^{hp}$ such that:

$$\mathcal{B}_{-}^{\sigma}(u_h,v) = \mathcal{F}_{-}^{\sigma}(v), \qquad \forall v \in \mathcal{V}^{hp}. \tag{3.20}$$

Remark 3 *Following Baker and Karakashian [5,6,16], we consider a variant of the SIPG method. Instead of using the formula (3.4), one may use:*

$$ac - bd = ac - ad + ad - bd = a(c-d) + (a-b)d \tag{3.21}$$

so that, by analogy:

$$\mathbf{n}\cdot(\nabla u)_i\,v_i - \mathbf{n}\cdot(\nabla u)_j\,v_j = \mathbf{n}\cdot(\nabla u)_i\,[v] + [\mathbf{n}\cdot\nabla u]\,v_j$$

and, since the fluxes, for $u \in H^2(\Omega)$, are continuous across the interelement boundaries, we have:

$$\int_{\gamma_{ij}} (\mathbf{n}\cdot\nabla u)_i\,v_i + (\mathbf{n}\cdot\nabla u)_j\,v_j\,\mathrm{ds} = \int_{\gamma_{ij}} \mathbf{n}\cdot(\nabla u)_i\,[v]\,\mathrm{ds}.$$

The new bilinear form for the boundary terms is now defined as:

$$I(u,v) = \int_{\Gamma_{int}\cup\Gamma_D} \mathbf{n}\cdot(\nabla u)_i\,[v]\,\mathrm{d}s$$

so that the new formulation reads: Find $u \in H^1(\Omega)\cap H^2(\mathcal{P}_h)$ *such that, for all* $v \in H^2(\mathcal{P}_h)$,

$$B(u,v) - I(u,v) - I(v,u) + J^{\sigma}(u,v) = F(v) - J_0(v) + J_0^{\sigma}(v). \tag{3.22}$$

We now see that we recover the SIPG method from the Baker-Karakashian formulation by replacing the term $\mathbf{n}\cdot(\nabla u)_i$ *by* $\langle \mathbf{n}\cdot\nabla u\rangle$. *It follows that all the properties associated with the SIPG method will also apply to the Baker-Karakashian formulation.*

3.2.3. Discontinuous *hp* Galerkin FE Method - DGM

The discontinuous Galerkin method by Baumann *et al.* [7,18] differs from the Global Element Method by just a sign. Indeed, by introducing the forms:

$$\begin{aligned} \mathcal{B}_+(u,v) &= B(u,v) - J(u,v) + J(v,u),\\ \mathcal{F}_+(v) &= F(v) + J_0(v), \end{aligned} \tag{3.23}$$

the DG formulation reads: Find u such that

$$\mathcal{B}_+(u,v) = \mathcal{F}_+(v), \qquad \forall v \in H^2(\mathcal{P}_h). \tag{3.24}$$

It is straightforward to show that the bilinear form is positive semidefinite.

The associated finite element version of the DG method consists then in finding $u_h \in \mathcal{V}^{hp}$ such that

$$\mathcal{B}_+(u_h,v) = \mathcal{F}_+(v), \qquad \forall v \in \mathcal{V}^{hp}. \tag{3.25}$$

3.2.4. Non-Symmetric Interior Penalty Galerkin Method - NIPG

This method was introduced by Rivière [19] and Süli, Schwab and Houston [22,15] and is inspired from the DG method with the addition of penalty terms. The new bilinear and linear forms read:

$$\begin{aligned} \mathcal{B}_+^{\sigma}(u,v) &= B(u,v) - J(u,v) + J(v,u) + J^{\sigma}(u,v),\\ \mathcal{F}_+^{\sigma}(v) &= F(v) + J_0(v) + J_0^{\sigma}(v), \end{aligned} \tag{3.26}$$

so that the problem to solve by the NIPG method becomes: Find u such that

$$\mathcal{B}_+^{\sigma}(u,v) = \mathcal{F}_+^{\sigma}(v), \qquad \forall v \in H^2(\mathcal{P}_h). \tag{3.27}$$

Once again, we may consider DG as a special case of NIPG with $\sigma = 0$.

The finite element problem corresponding to the NIPG formulation (3.27) is to find $u_h \in \mathcal{V}^{hp}$ such that

$$\mathcal{B}_+^\sigma(u_h, v) = \mathcal{F}_+^\sigma(v), \qquad \forall v \in \mathcal{V}^{hp}. \tag{3.28}$$

The four methods presented thus far are all very similar, except for a plus or minus sign in front of the term $J(v, u)$ and the addition of a penalty term $J^\sigma(u, v)$ or not. We shall now see how these changes modify the properties of the respective formulations.

3.3. Equivalence of Strong and Weak Problems

We shall show the equivalence of the strong and weak formulations only with respect to the Global Element method. The results are identical for the other formulations, namely the SIPG, DG and NIPG methods. Existence of solutions of the discontinuous formulations is then somewhat guaranteed. However, we emphasize here that Theorem 3.1 does not infer anything about the uniqueness of the solutions. This question still remains an open issue.

Theorem 3.1 (GE Method) *Let $u \in C^2(\overline{\Omega})$ be the solution of Problem* (3.1)-(3.2). *Then u satisfies the weak formulation* (3.16). *Conversely, if $u \in H^1(\Omega) \cap H^2(\mathcal{P}_h)$ is a solution of* (3.16) *then u satisfies the partial differential equation* (3.1) *and boundary conditions* (3.2).

Proof: The first part of the theorem has been proved along with the derivation of the Global Element formulation, since (3.9) is satisfied when $u \in C^2(\overline{\Omega})$.

The converse follows the proof given in Rivière [19]. Let $\mathcal{D}(K) \subset H^2(K)$ be the space of infinitely differentiable functions with compact support on element K and let $v \in \mathcal{D}(K)$. Then (3.16) gives:

$$\int_K (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x = \int_K fv\,\mathrm{d}x$$

which implies, after integration by parts and since v is arbitrary in $\mathcal{D}(K)$, that

$$-\Delta u + cu = f, \qquad \text{a.e. in } K. \tag{3.29}$$

Next, we consider an interior edge γ_{ij} shared by the elements K_i and K_j. Let v be a function in $H_0^2(K_i \cup K_j) \subset H^2(K_i) \times H^2(K_j)$, extended by zero outside. Then the boundary terms $J(u, v)$ and $J(v, u)$ vanish, because $[u] = [v] = 0$ on γ_{ij}, and the weak formulation (3.16) reduces to

$$\int_{K_i \cup K_j} (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x = \int_{K_i \cup K_j} fv\,\mathrm{d}x \tag{3.30}$$

On the other hand, multiplying (3.29) by v, integrating on K_i and K_j and using Green's formula, we have:

$$\int_{K_i} (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x - \int_{\gamma_{ij}} (\mathbf{n} \cdot \nabla u)_i v\,\mathrm{d}s = \int_{K_i} fv\,\mathrm{d}x,$$

$$\int_{K_j} (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x - \int_{\gamma_{ij}} (\mathbf{n} \cdot \nabla u)_j v\,\mathrm{d}s = \int_{K_j} fv\,\mathrm{d}x,$$

so that

$$\int_{K_i \cup K_j} (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x - \int_{\gamma_{ij}} [\mathbf{n} \cdot \nabla u]\, v\,\mathrm{d}s = \int_{K_i \cup K_j} fv\,\mathrm{d}x. \tag{3.31}$$

Comparing (3.30) and (3.31), one observes that:

$$\int_{\gamma_{ij}} [\mathbf{n} \cdot \nabla u]\, v\,\mathrm{d}s = 0, \qquad \forall v \in H_0^2(K_i \cup K_j).$$

Then, $[\mathbf{n} \cdot \nabla u] = 0$ for all element edges γ_{ij}, which implies $\nabla u \in H(div, \Omega)$. This allows us to conclude that u satisfies Poisson Equation globally on Ω, i.e.

$$-\Delta u + cu = f, \qquad \text{a.e. in } \Omega. \tag{3.32}$$

To recover the Dirichlet boundary conditions, we now consider a function $v \in H_0^1(\Omega) \cap H^2(\Omega)$, so that integrating (3.32) provides:

$$\int_\Omega (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x = \int_\Omega fv\,\mathrm{d}x,$$

whereas (3.16) yields:

$$\int_\Omega (\nabla u \cdot \nabla v + cuv)\,\mathrm{d}x - \int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\, u\,\mathrm{d}s = \int_\Omega fv\,\mathrm{d}x - \int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\, u_0\,\mathrm{d}s.$$

Substracting both equations, we obtain:

$$\int_{\Gamma_D} (\mathbf{n} \cdot \nabla v)\,(u - u_0)\,\mathrm{d}s = 0, \qquad \forall v \in H_0^1(\Omega) \cap H^2(\Omega),$$

and conclude that $u = u_0$ on Γ_D.

In the same way, choosing $v \in H^2(\Omega) \subset H^2(\mathcal{T}_h)$ such that $v = 0$ on Γ_D, we get:

$$\int_{\Gamma_N} (\mathbf{n} \cdot \nabla u - g)\, v\,\mathrm{d}s = 0,$$

so that $\mathbf{n} \cdot \nabla u = g$ on Γ_N. □

Remark 4 *When c is zero, $C^2(\overline{\Omega})$ can be replaced in Theorem 3.1 by $H^1(\Omega) \cap H^2(\mathcal{T}_h)$ since $\nabla u \in H(div, \Omega)$.*

3.4. Properties of the Bilinear Forms

3.4.1. Mesh-dependent norms

We now introduce norms associated with the bilinear forms:

1. Energy Norm:

$$\|v\|_{e,\mathcal{P}_h}^2 = B(v,v) = \sum_{K\in\mathcal{P}_h} \|v\|_{e,K}^2 = \sum_{K\in\mathcal{P}_h} \left(\|\nabla v\|_{0,K}^2 + c\|v\|_{0,K}^2\right) \tag{3.33}$$

2. Norm proposed by Süli *et al.* in [22,15]:

$$\|v\|_{\mathcal{P}_h}^2 = B(v,v) + J^\sigma(v,v) = \|v\|_{e,\mathcal{P}_h}^2 + \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[v]^2\,\mathrm{d}s \tag{3.34}$$

3. Norm proposed by Baumann *et al.*in [7,17,18] and by Baker and Karakashian in [6]:

$$|||v|||_{\mathcal{P}_h}^2 = \|v\|_{\mathcal{P}_h}^2 + \int_{\Gamma_{int}\cup\Gamma_D} \frac{1}{\sigma}\,\langle\mathbf{n}\cdot\nabla v\rangle^2\,\mathrm{d}s \tag{3.35}$$

We note that the energy norm becomes a seminorm when c is zero.

3.4.2. Continuity of the bilinear forms

We shall show now that the bilinear forms $\mathcal{B}_\pm(\cdot,\cdot)$ and $\mathcal{B}_\pm^\sigma(\cdot,\cdot)$ are continuous on $H^2(\mathcal{P}_h)$ with respect to the norm $|||\cdot|||_{\mathcal{P}_h}$ defined in (3.35). Unfortunately, we are unable to show continuity with respect to the other two norms (3.33) and (3.34).

Theorem 3.2 (GEM and DGM) *Let $\mathcal{B}_\pm(\cdot,\cdot)$ be the bilinear form defined either in (3.15) or in (3.23) . Then,*

$$|\mathcal{B}_\pm(u,v)| \le |||u|||_{\mathcal{P}_h}\,|||v|||_{\mathcal{P}_h}, \qquad \forall u,\forall v\in H^2(\mathcal{P}_h). \tag{3.36}$$

Proof:

First note that:

$$\begin{aligned}|\mathcal{B}_\pm(u,v)| &= |B(u,v) - J(u,v) \pm J(v,u)| \\ &\le |B(u,v)| + |J(u,v)| + |J(v,u)|\end{aligned}$$

It is clear that

$$|B(u,v)| \le \sum_{K\in\mathcal{P}_h} \int_K |\nabla u\cdot\nabla v + cuv|\,\mathrm{d}x \le \|u\|_{e,\mathcal{P}_h}\|v\|_{e,\mathcal{P}_h}$$

The first boundary term gives:

$$
\begin{aligned}
|J(u,v)| &\le \int_{\Gamma_{int}\cup\Gamma_D} |\langle \mathbf{n}\cdot\nabla u\rangle\,[v]|\,\mathrm{ds}\\
&\le \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla u\rangle^2\,\mathrm{ds}}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[v]^2\,\mathrm{ds}}.
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
|J(v,u)| &\le \int_{\Gamma_{int}\cup\Gamma_D} |\langle \mathbf{n}\cdot\nabla v\rangle\,[u]|\,\mathrm{ds}\\
&\le \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[u]^2\,\mathrm{ds}}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla v\rangle^2\,\mathrm{ds}}.
\end{aligned}
$$

In consequence, we have, using the discrete Schwarz inequality (A.1):

$$
\begin{aligned}
|\mathcal{B}_{\pm}(u,v)| \le\; & \|u\|_{e,\mathcal{T}_h}\|v\|_{e,\mathcal{T}_h}\\
&+ \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla u\rangle^2\,\mathrm{ds}}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[v]^2\,\mathrm{ds}}\\
&+ \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[u]^2\,\mathrm{ds}}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla v\rangle^2\,\mathrm{ds}}.\\
\le\; & \sqrt{\|u\|^2_{e,\mathcal{T}_h} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[u]^2\,\mathrm{ds} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla u\rangle^2\,\mathrm{ds}}\\
&\times \sqrt{\|v\|^2_{e,\mathcal{T}_h} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,\,[v]^2\,\mathrm{ds} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\,\langle \mathbf{n}\cdot\nabla v\rangle^2\,\mathrm{ds}}\\
\le\; & |||u|||_{\mathcal{T}_h}\,|||v|||_{\mathcal{T}_h},
\end{aligned}
$$

which completes the proof. □

Theorem 3.3 (SIPG and NIPG Methods) *Let $\mathcal{B}^{\sigma}_{\pm}(\cdot,\cdot)$ be the bilinear form defined either in (3.18) or in (3.26). Then,*

$$
|\mathcal{B}^{\sigma}_{\pm}(u,v)| \le C\,|||u|||_{\mathcal{T}_h}\,|||v|||_{\mathcal{T}_h}, \qquad \forall u, \forall v \in H^2(\mathcal{T}_h). \tag{3.37}
$$

where C is a constant, $C \le 2$.

Proof:

As before we have:

$$
\begin{aligned}
|\mathcal{B}^{\sigma}_{\pm}(u,v)| &= |B(u,v) - J(u,v) \pm J(v,u) + J^{\sigma}(u,v)|\\
&\le |B(u,v)| + |J(u,v)| + |J(v,u)| + |J^{\sigma}(u,v)|\\
&\le |||u|||_{\mathcal{T}_h}\,|||v|||_{\mathcal{T}_h} + |J^{\sigma}(u,v)|.
\end{aligned}
$$

And

$$|J^{\sigma}(u,v)| \leq \int_{\Gamma_{int}\cup\Gamma_D} |\sigma\,[u]\,[v]|\,\mathrm{d}s \leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[u]^2\,\mathrm{d}s}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[v]^2\,\mathrm{d}s}.$$

Therefore, making use again of the discrete Schwarz inequality (A.1), we obtain:

$$\begin{aligned}
|\mathcal{B}^{\sigma}_{\pm}(u,v)| &\leq |||u|||_{\mathcal{P}_h}\,|||v|||_{\mathcal{P}_h} + \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[u]^2\,\mathrm{d}s}\,\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[v]^2\,\mathrm{d}s}. \\
&\leq \sqrt{|||u|||^2_{\mathcal{P}_h} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[u]^2\,\mathrm{d}s}\sqrt{|||v|||^2_{\mathcal{P}_h} + \int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[v]^2\,\mathrm{d}s} \\
&\leq \sqrt{2\,|||u|||^2_{\mathcal{P}_h}}\sqrt{2\,|||v|||^2_{\mathcal{P}_h}} \\
&\leq 2\,|||u|||_{\mathcal{P}_h}\,|||v|||_{\mathcal{P}_h},
\end{aligned}$$

and we see that C is at most equal to 2. □

3.4.3. Coercivity of the bilinear forms in the discrete spaces

Here we wish to show that the bilinear forms $\mathcal{B}_{\pm}(\cdot,\cdot)$ and $\mathcal{B}^{\sigma}_{\pm}(\cdot,\cdot)$ are coercive in $H^2(\mathcal{P}_h)$ with respect to the norm $|||\cdot|||_{\mathcal{P}_h}$ in order to be able to apply classical theorems for existence and uniqueness of solutions of the discontinuous methods. Unfortunately, to date, we are able to prove coercivity only in the discrete discontinuous spaces $\mathcal{V}^{hp}$, and then, only for the SIPG and NIPG formulations .

Theorem 3.4 (NIPG Method) *Let $\sigma = \kappa p^2/h$, κ being a positive number. Then, for all $\kappa > 0$, there exists a positive constant, $\alpha > 0$, such that:*

$$\mathcal{B}^{\sigma}_{+}(z,z) \geq \alpha\,|||z|||^2_{\mathcal{P}_h}, \qquad \forall z \in \mathcal{V}^{hp}. \tag{3.38}$$

Here α is independent of h and p.

Proof: Let α be an arbitrary real number and choose a $z \in \mathcal{V}^{hp}$. Then

$$\begin{aligned}
&\mathcal{B}^{\sigma}_{+}(z,z) - \alpha\,|||z|||^2_{\mathcal{P}_h} \\
&\qquad = (1-\alpha)\,B(z,z) + (1-\alpha)\,J^{\sigma}(z,z) - \alpha\int_{\Gamma_{int}\cup\Gamma_D} \frac{1}{\sigma}\,\langle \mathbf{n}\cdot\nabla z\rangle^2\,\mathrm{d}s
\end{aligned}$$

Since $\langle \mathbf{n}\cdot\nabla z\rangle$ is the average of the flux at the interface of two elements K_i and K_j, the corresponding integral can be split into two integrals with integrands $(\mathbf{n}\cdot\nabla z)_i/\sigma$ and $(\mathbf{n}\cdot\nabla z)_j/\sigma$, each one associated with the elements K_i or K_j respectively. Therefore, let $\gamma \subset \Gamma_{int}\cup\Gamma_D$ and consider the integral associated with the element

K. Using the trace inequality (A.3) and the inverse property (A.7), we have

$$\begin{aligned}
\int_\gamma \frac{1}{\sigma}(\mathbf{n}\cdot\nabla z)^2\,ds &\le \frac{1}{\sigma}\|\nabla z\|_{0,\gamma}^2 \\
&\le \frac{C}{\sigma}\left(\frac{1}{h_K}\|\nabla z\|_{0,K}^2 + \|\nabla z\|_{0,K}\|\nabla^2 z\|_{0,K}\right) \\
&\le \frac{C}{\sigma}\left(\frac{1}{h_K} + C_0\frac{p_K^2}{h_K}\right)\|\nabla z\|_{0,K}^2 \\
&\le \frac{C}{\sigma}\frac{p_K^2}{h_K}\|\nabla z\|_{0,K}^2,
\end{aligned}$$

so that, selecting σ to be equal to $\kappa p_K^2/h_K$, we obtain:

$$-\int_\Gamma \frac{1}{\sigma}(\mathbf{n}\cdot\nabla z)^2\,ds \ge -\frac{C}{\kappa}\|\nabla z\|_{0,K}^2.$$

Note that, when the mesh size h_{K_i} and h_{K_j} and the polynomial degrees p_{K_i} and p_{K_j} are different from each other in the two elements K_i and K_j sharing the edge γ_{ij}, we actually choose σ as

$$\sigma = \kappa\frac{\max(p_{K_i}^2, p_{K_j}^2)}{\min(h_{K_i}, h_{K_j})},$$

so that:

$$\begin{aligned}
\int_{\gamma_{ij}} \frac{1}{\sigma}(\mathbf{n}\cdot\nabla z)_i^2\,ds &\le \frac{C}{\sigma}\frac{p_{K_i}^2}{h_{K_i}}\|\nabla z\|_{0,K_i}^2 \\
&\le \frac{C}{\kappa}\frac{\min(h_{K_i}, h_{K_j})}{\max(p_{K_i}^2, p_{K_j}^2)}\frac{p_{K_i}^2}{h_{K_i}}\|\nabla z\|_{0,K_i}^2 \\
&\le \frac{C}{\kappa}\|\nabla z\|_{0,K_i}^2.
\end{aligned}$$

It then follows that:

$$\mathcal{B}_+^\sigma(z,z) - \alpha\,|||z|||_{\mathcal{P}_h}^2 \ge (1-\alpha-\alpha C/\kappa)\,B(z,z) + (1-\alpha)\,J^\sigma(z,z).$$

Therefore, we certainly can pick a value of α such that

$$0 < \alpha \le \frac{1}{1+C/\kappa}$$

for which the bilinear form $\mathcal{B}_+^\sigma(\cdot,\cdot)$ is coercive in $\mathcal{V}^{hp}$, for all $\kappa > 0$. $\square$

Theorem 3.5 (SIPG Method) *Let $\sigma = \kappa p^2/h$, κ being a positive number. Then, for $\kappa > \kappa_0$, there exists a positive constant α independent of h and p, $\alpha > 0$, such that:*

$$\mathcal{B}^{\sigma}_{-}(z,z) \geq \alpha \,|||z|||^2_{\mathcal{T}_h}, \qquad \forall z \in \mathcal{V}^{hp}. \tag{3.39}$$

Proof: Let α be an arbitrary real number and choose $z \in \mathcal{V}^{hp}$. Then

$$\mathcal{B}^{\sigma}_{-}(z,z) - \alpha \,|||z|||^2_{\mathcal{T}_h} = (1-\alpha)\, B(z,z) + (1-\alpha)\, J^{\sigma}(z,z)$$
$$- 2\int_{\Gamma_{int}\cup\Gamma_D} \langle \mathbf{n}\cdot\nabla z\rangle [z]\, ds - \alpha \int_{\Gamma_{int}\cup\Gamma_D} \frac{1}{\sigma}\langle \mathbf{n}\cdot\nabla z\rangle^2\, ds$$

There exists a positive number ε such that for every edge $\gamma \in \Gamma_{int} \cup \Gamma_D$:

$$2\int_{\gamma} \langle \mathbf{n}\cdot\nabla z\rangle [z]\, ds \leq 2\sqrt{\int_{\gamma} \sigma^{-1}\langle \mathbf{n}\cdot\nabla z\rangle^2\, ds}\sqrt{\int_{\gamma} \sigma\, [z]^2\, ds}$$
$$\leq \varepsilon \int_{\gamma} \frac{1}{\sigma}\langle \mathbf{n}\cdot\nabla z\rangle^2\, ds + \frac{1}{\varepsilon}\int_{\gamma} \sigma\, [z]^2\, ds$$

which yields, using the result in the previous proof:

$$\mathcal{B}^{\sigma}_{-}(z,z) - \alpha \,|||z|||^2_{\mathcal{T}_h} \geq \left(1 - \alpha - (\alpha+\varepsilon)\frac{C}{\kappa}\right) B(z,z) + \left(1 - \alpha - \frac{1}{\varepsilon}\right) J^{\sigma}(z,z).$$

In order to prove coercivity, we want to find $\alpha > 0$ such that both factors in the inequality are positive, in other words:

$$\left(1 - \alpha - (\alpha+\varepsilon)\frac{C}{\kappa}\right) > 0 \qquad \text{and} \qquad \left(1 - \alpha - \frac{1}{\varepsilon}\right) > 0.$$

The second inequality requires that:

$$0 < \alpha \leq 1 - \frac{1}{\varepsilon}$$

which means that

$$\varepsilon > 1.$$

On the other hand the first inequality requires that:

$$0 < \alpha \leq \frac{1-\varepsilon C/\kappa}{1+C/\kappa} \leq \frac{1-C/\kappa}{1+C/\kappa} \leq \frac{\kappa - C}{\kappa + C}$$

This completes the proof by taking κ sufficiently large, namely $\kappa \geq \kappa_0$ (where for instance $\kappa_0 > C$.) $\square$

Remark 5 *We note that $\mathcal{B}_{+}^{\sigma}(\cdot,\cdot)$ (for NIPG Method) is coercive in $H^2(\mathcal{T}_h)$ with respect to the norm $\|\cdot\|_{\mathcal{T}_h}$. Indeed, for all $v \in H^2(\mathcal{T}_h)$,*

$$\mathcal{B}_{+}^{\sigma}(v,v) = B(v,v) - J(v,v) + J(v,v) + J^{\sigma}(v,v) = \|v\|_{\mathcal{T}_h}^2. \tag{3.40}$$

It is also straightforward to show that $\mathcal{B}_{+}(\cdot,\cdot)$ (for DGM) is coercive in $H^2(\mathcal{T}_h)$ with respect to the energy norm $\|\cdot\|_{e,\mathcal{T}_h}$:

$$\mathcal{B}_{+}(v,v) = B(v,v) - J(v,v) + J(v,v) = B(v,v) = \|v\|_{e,\mathcal{T}_h}^2. \tag{3.41}$$

These results will be crucial in deriving a priori error estimates in the next section.

4. *A Priori* Error Estimates

4.1. SIPG and NIPG Methods

Theorem 4.1 *Let $u \in H^1(\Omega) \cap H^s(\mathcal{T}_h)$, $s \geq 2$, be a solution of (3.18) (SIPG) or (3.26) (NIPG) and u_h be the discrete discontinuous solution of*

$$\mathcal{B}_{\pm}^{\sigma}(u_h, v) = \mathcal{F}(v), \qquad \forall v \in \mathcal{V}^{hp}. \tag{4.1}$$

Then, choosing $\sigma = \kappa p^2/h$, ($\kappa > 0$ for NIPG and $\kappa \geq \kappa_0$ for SIPG), the numerical error $e = u - u_h$ satisfies:

$$\|e\|_{e,\mathcal{T}_h} \leq C\frac{h^{\mu-1}}{p^{s-3/2}}\|u\|_s \tag{4.2}$$

where $\mu = \min(p+1, s)$ and $p \geq 1$.

4.1.1. Proof of Theorem 4.1 for SIPG and NIPG

First, by definition of the norms, we note that $\|e\|_{e,\mathcal{T}_h} \leq |||e|||_{\mathcal{T}_h}$. In other words, it suffices here to estimate the error with respect to the norm $|||\cdot|||_{\mathcal{T}_h}$. The proof is inspired by [5,6,16] where the authors have derived the rate of convergence in h only for the SIPG method of the (3.22) form. Here we extend their results to the NIPG formulation as well and also show for both methods the rate of convergence in p.

Proof: Let z_p be an interpolant of u in $\mathcal{V}^{hp}$. We shall use the notation $\eta = u - z_p$ and $\xi = u_h - z_p$ so that $e = u - u_h = \eta - \xi$. Applying the triangle inequality, we have:

$$|||e|||_{\mathcal{T}_h} = |||u - u_h|||_{\mathcal{T}_h} = |||\eta - \xi|||_{\mathcal{T}_h} \leq |||\eta|||_{\mathcal{T}_h} + |||\xi|||_{\mathcal{T}_h}.$$

From the coercivity of the bilinear form $\mathcal{B}^{\sigma}_{\pm}(\cdot,\cdot)$, since $\xi \in \mathcal{V}^{hp}$, we have

$$|||\xi|||^2_{\mathcal{T}_h} \leq C\mathcal{B}^{\sigma}_{\pm}(\xi,\xi),$$

and from the "orthogonality" property $\mathcal{B}^{\sigma}_{\pm}(u-u_h,v)=0, \forall v \in \mathcal{V}^{hp}$, we get

$$\mathcal{B}^{\sigma}_{\pm}(\xi,\xi)=\mathcal{B}^{\sigma}_{\pm}(\eta,\xi), \qquad \forall v \in \mathcal{V}^{hp}.$$

Using the continuity of $\mathcal{B}^{\sigma}_{\pm}(\cdot,\cdot)$, we know that

$$\mathcal{B}^{\sigma}_{\pm}(\eta,\xi) \leq C\,|||\eta|||_{\mathcal{T}_h}\,|||\xi|||_{\mathcal{T}_h},$$

which implies

$$|||\xi|||_{\mathcal{T}_h} \leq C\,|||\eta|||_{\mathcal{T}_h}.$$

Finally, we have

$$|||e|||_{\mathcal{T}_h} \leq |||\eta|||_{\mathcal{T}_h} + |||\xi|||_{\mathcal{T}_h} \leq C\,|||\eta|||_{\mathcal{T}_h}.$$

We recall here that C is a generic constant independent of h and p which takes different values at different places.

We now choose the interpolant z_p as defined in Lemma A.7. Then:

$$|||\eta|||^2_{\mathcal{T}_h} = \sum_{K\in\mathcal{T}_h}\int_K\left(|\nabla\eta|^2+c\eta^2\right)\mathrm{d}x+\int_{\Gamma_{int}\cup\Gamma_D}\frac{1}{\sigma}\langle\mathbf{n}\cdot\nabla\eta\rangle^2\,\mathrm{d}s+\int_{\Gamma_{int}\cup\Gamma_D}\sigma\,[\eta]^2\,\mathrm{d}s \tag{4.3}$$

The integrals in the leading term are estimated as, using (A.8):

$$\int_K|\nabla\eta|^2\,\mathrm{d}x \leq C\left(\frac{h_K^{\mu-1}}{p_K^{s-1}}\right)^2\|u\|^2_{s,K}, \qquad s\geq 1,$$

$$\int_K c\eta^2\,\mathrm{d}x \leq cC\left(\frac{h_K^{\mu}}{p_K^{s}}\right)^2\|u\|^2_{s,K}, \qquad s\geq 0,$$

so that

$$\int_K\left(|\nabla\eta|^2+c\eta^2\right)\mathrm{d}x \leq C\frac{h_K^{2\mu-2}}{p_K^{2s-2}}\|u\|^2_{s,K}, \qquad s\geq 1.$$

Let γ_{ij} denote an interior edge shared by the elements K_i and K_j. Then, using the inequality $(a+b)^2 \leq 2a^2+2b^2$, we observe that

$$\int_{\gamma_{ij}}\frac{1}{\sigma}\langle\mathbf{n}\cdot\nabla\eta\rangle^2\,\mathrm{d}s \leq \frac{1}{2}\int_{\gamma_{ij}}\frac{1}{\sigma}(\mathbf{n}\cdot(\nabla\eta)_i)^2\,\mathrm{d}s+\frac{1}{2}\int_{\gamma_{ij}}\frac{1}{\sigma}\left(\mathbf{n}\cdot(\nabla\eta)_j\right)^2\mathrm{d}s.$$

In other words, in splitting the second integrals on the right hand side of (4.3) as above, we actually associate with each $\gamma \in \mathcal{E}_{h,int} \cup \mathcal{E}_{h,D}$ an element K, such that

$$
\begin{aligned}
\int_\gamma \frac{1}{\sigma} (\mathbf{n} \cdot \nabla \eta)^2 \, ds &\le \frac{1}{\sigma} \|\nabla \eta\|_{0,\gamma}^2 \\
&\le \frac{C}{\sigma} \left(\frac{1}{h_K} \|\nabla \eta\|_{0,K}^2 + \|\nabla \eta\|_{0,K} \|\nabla^2 \eta\|_{0,K} \right) \\
&\le \frac{C}{\sigma} \left(\frac{1}{h_K} \|\eta\|_{1,K}^2 + \|\eta\|_{1,K} \|\eta\|_{2,K} \right) \\
&\le \frac{C}{\sigma} \left(\frac{1}{h_K} \frac{h_K^{2\mu-2}}{p_K^{2s-2}} + \frac{h_K^{\mu-1}}{p_K^{s-1}} \frac{h_K^{\mu-2}}{p_K^{s-2}} \right) \|u\|_{s,K}^2 \\
&\le \frac{C}{\sigma} \left(\frac{h_K^{2\mu-3}}{p_K^{2s-2}} + \frac{h_K^{2\mu-3}}{p_K^{2s-3}} \right) \|u\|_{s,K}^2 \\
&\le \frac{C}{\sigma} \frac{h_K^{2\mu-3}}{p_K^{2s-3}} \|u\|_{s,K}^2 \\
&\le \frac{C}{\kappa} \frac{h_K^{2\mu-2}}{p_K^{2s-1}} \|u\|_{s,K}^2, \qquad s \ge 2.
\end{aligned}
$$

Again, for an interior edge γ_{ij} shared by K_i and K_j, using $(a-b)^2 \le 2a^2 + 2b^2$, we have:

$$
\int_{\gamma_{ij}} \sigma [\eta]^2 \, ds = \int_{\gamma_{ij}} \sigma \left(\eta_i - \eta_j\right)^2 ds \le 2 \int_{\gamma_{ij}} \sigma (\eta_i)^2 \, ds + 2 \int_{\gamma_{ij}} \sigma \left(\eta_j\right)^2 ds
$$

This means that the edge integrals making the third term of (4.3) are bounded by:

$$
\int_\gamma \sigma (\eta)^2 \, ds \le C\sigma \frac{h_K^{2\mu-1}}{p_K^{2s-1}} \|u\|_{s,K}^2 \le C\kappa \frac{h_K^{2\mu-2}}{p_K^{2s-3}} \|u\|_{s,K}^2
$$

In combining the above results, we thus obtain

$$
\begin{aligned}
|\!|\!| e |\!|\!|_{\mathcal{P}_h} \le C |\!|\!| \eta |\!|\!|_{\mathcal{P}_h} &\le C \sum_{K \in \mathcal{P}_h} \left\{ \frac{h_K^{2\mu-2}}{p_K^{2s-2}} + \frac{h_K^{2\mu-2}}{p_K^{2s-1}} + \frac{h_K^{2\mu-2}}{p_K^{2s-3}} \right\}^{1/2} \|u\|_{s,K} \\
&\le C \sum_{K \in \mathcal{P}_h} \frac{h_K^{\mu-1}}{p_K^{s-3/2}} \|u\|_{s,K} \\
&\le C \frac{h^{\mu-1}}{p^{s-3/2}} \|u\|_s
\end{aligned}
$$

which is the expected *a priori* error estimate. □

4.1.2. Alternative Proof of Theorem 4.1 for NIPG

Alternatively, we present a second proof of Theorem 4.1 for the NIPG method only as it is based on the nonsymmetry of the formulation. The proof is inspired by the one found in [22]. However, our rate of convergence with respect to p was improved from $(s-2)$ to $(s-3/2)$ using the interpolation estimates of Lemma A.7. Later, the same authors proposed in [15] a comparable version of the proof with $(s-3/2)$ as the rate of convergence.

Proof: Once again, z_p is the interpolant of u in $\mathcal{V}^{hp}$ as defined in Lemma A.7. and we denote $\eta = u - z_p$ and $\xi = u_h - z_p$ as before. Then,

$$||e||_{e,\mathcal{T}_h} \leq ||e||_{\mathcal{T}_h} = ||u - u_h||_{\mathcal{T}_h} = ||\eta - \xi||_{\mathcal{T}_h} \leq ||\eta||_{\mathcal{T}_h} + ||\xi||_{\mathcal{T}_h}.$$

Moreover, from the definition of $\mathcal{B}^\sigma_+(\cdot,\cdot)$ and the norm $||\cdot||_{e,\mathcal{T}_h}$ (see (3.40)) and the "orthogonality" relation, we have:

$$||\xi||^2_{\mathcal{T}_h} = \mathcal{B}^\sigma_+(\xi,\xi) = \mathcal{B}^\sigma_+(\eta,\xi).$$

The goal is now to bound $\mathcal{B}^\sigma_+(\eta,\xi)$ in terms of $||\xi||_{\mathcal{T}_h}$. Recall that:

$$\begin{aligned}\mathcal{B}^\sigma_+(\eta,\xi) &= B(\eta,\xi) + J^\sigma(\eta,\xi) - J(\eta,\xi) + J(\xi,\eta)\\ &\leq |B(\eta,\xi)| + |J^\sigma(\eta,\xi)| + |J(\eta,\xi)| + |J(\xi,\eta)|\end{aligned}$$

The first term on the right hand side of the equation above gives:

$$|B(\eta,\xi)| \leq \sum_{K\in\mathcal{T}_h} \int_K |\nabla\eta\cdot\nabla\xi + c\eta\xi|\,\mathrm{d}x \leq ||\eta||_{e,\mathcal{T}_h}||\xi||_{e,\mathcal{T}_h} \leq ||\eta||_{\mathcal{T}_h}||\xi||_{\mathcal{T}_h}.$$

The term $|J^\sigma(\eta,\xi)|$ is bounded by:

$$\begin{aligned}|J^\sigma(\eta,\xi)| &\leq \int_{\Gamma_{int}\cup\Gamma_D} |\sigma\,[\eta]\,[\xi]|\,\mathrm{d}s\\ &\leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[\eta]^2\,\mathrm{d}s}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[\xi]^2\,\mathrm{d}s}\\ &\leq ||\eta||_{\mathcal{T}_h}||\xi||_{\mathcal{T}_h},\end{aligned}$$

whereas we have for the third term:

$$\begin{aligned}|J(\eta,\xi)| &\leq \int_{\Gamma_{int}\cup\Gamma_D} |\langle\mathbf{n}\cdot\nabla\eta\rangle\,[\xi]|\,\mathrm{d}s\\ &\leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\langle\mathbf{n}\cdot\nabla\eta\rangle^2\,\mathrm{d}s}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma\,[\xi]^2\,\mathrm{d}s}\\ &\leq ||\xi||_{\mathcal{T}_h}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1}\langle\mathbf{n}\cdot\nabla\eta\rangle^2\,\mathrm{d}s}.\end{aligned}$$

Likewise, $J(\xi,\eta)$ is bounded by:

$$|J(\xi,\eta)| \leq \|\eta\|_{\mathcal{T}_h} \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \sigma^{-1} \langle \mathbf{n}\cdot\nabla\xi\rangle^2 \, \mathrm{d}s}$$

Using again the trace inequality (A.3) and the inverse property (A.7), it is shown that:

$$\int_\gamma \frac{1}{\sigma}(\mathbf{n}\cdot\nabla\xi)^2 \, \mathrm{d}s \leq \frac{C}{\sigma}\frac{p_K^2}{h_K} \|\nabla\xi\|_{0,K}^2$$

In other words, using $\sigma = \kappa p_K^2/h_K$

$$|J(\xi,\eta)| \leq C\|\eta\|_{\mathcal{T}_h}\|\xi\|_{\mathcal{T}_h}$$

Combining the above results, we have:

$$\|\xi\|_{\mathcal{T}_h} \leq C\left(\|\eta\|_{\mathcal{T}_h} + \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \frac{1}{\sigma}\langle \mathbf{n}\cdot\nabla\eta\rangle^2 \, \mathrm{d}s} \right) \leq C \,|\!|\!|\eta|\!|\!|_{\mathcal{T}_h}$$

so that:

$$\|e\|_{e,\mathcal{T}_h} \leq \|e\|_{\mathcal{T}_h} \leq \|\eta\|_{\mathcal{T}_h} + \|\xi\|_{\mathcal{T}_h} \leq \|\eta\|_{\mathcal{T}_h} + C\,|\!|\!|\eta|\!|\!|_{\mathcal{T}_h} \leq C\,|\!|\!|\eta|\!|\!|_{\mathcal{T}_h}.$$

We conclude the proof by employing the estimate on $|\!|\!|\eta|\!|\!|_{\mathcal{T}_h}$ shown in the previous proof. □

4.2. DG Method

We recall that the DG formulation proposed in [7,18] is deduced from the NIPG method by simply setting the penalty parameter σ to zero. However, unlike NIPG, continuity and coercivity of the bilinear form $\mathcal{B}_+(\cdot,\cdot)$ cannot be proved simultaneously using the same norm. At best it is shown that:

$$B(v,v) = \|v\|_{e,\mathcal{T}_h}^2, \qquad \forall v \in H^2(\mathcal{T}_h),$$

and that:

$$\mathcal{B}_+(u,v) \leq |\!|\!|u|\!|\!|_{\mathcal{T}_h} |\!|\!|v|\!|\!|_{\mathcal{T}_h}, \qquad \forall u,v \in H^2(\mathcal{T}_h).$$

The main issue in finding *a priori* error estimates for the error $e = u - u_h$ in the numerical approximation u_h of the DG problem consists in deriving an upper bound on:

$$\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} [\xi]^2 \, \mathrm{d}s}$$

with respect to the norm $\|\xi\|_{e,\mathcal{T}_h}$ when $c = 0$. This integral does indeed appear when bounding the term $J(\eta, \xi)$, i.e.

$$|J(\eta,\xi)| \leq \int_{\Gamma_{int}\cup\Gamma_D} |\langle \mathbf{n}\cdot\nabla\eta\rangle\,[\xi]|\,\mathrm{d}s \leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \langle \mathbf{n}\cdot\nabla\eta\rangle^2\,\mathrm{d}s}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} [\xi]^2\,\mathrm{d}s}.$$

We present below two approaches, by treating separately the case when c is zero and the case when c is nonzero.

4.2.1. *A priori* error estimate when c is nonzero

We find it instructive to analyze the special case in which c is strictly greater than zero. In this case, we still can use the methodology presented earlier for the NIPG method. However, we shall see that the rate of convergence with respect to the mesh size becomes suboptimal as stated in the following theorem.

Theorem 4.2 *Let $u \in H^1(\Omega)\cap H^s(\mathcal{T}_h)$, $s \geq 2$, be a solution of (3.23) with $c > 0$ and u_h be the discrete discontinuous solution of (3.24). Then, the numerical error $e = u - u_h$ satisfies:*

$$\|e\|_{e,\mathcal{T}_h} \leq C\frac{h^{\mu-2}}{p^{s-3/2}}\|u\|_s \tag{4.4}$$

where $\mu = \min(p+1, s)$ and $p \geq 1$.

Proof: Using the same procedure and notation as before, we have:

$$\|e\|_{e,\mathcal{T}_h} = \|u-u_h\|_{e,\mathcal{T}_h} = \|\eta-\xi\|_{e,\mathcal{T}_h} \leq \|\eta\|_{e,\mathcal{T}_h} + \|\xi\|_{e,\mathcal{T}_h}.$$

Moreover, from the definition of $\mathcal{B}_+(\cdot,\cdot)$ (see (3.41)) and the "orthogonality" relation, we further show that:

$$\begin{aligned}
\|\xi\|^2_{e,\mathcal{T}_h} &= \mathcal{B}_+(\xi,\xi)\\
&= \mathcal{B}_+(\eta,\xi)\\
&= B(\eta,\xi) - J(\eta,\xi) + J(\xi,\eta)\\
&\leq |B(\eta,\xi)| + |J(\eta,\xi)| + |J(\xi,\eta)|
\end{aligned}$$

We now consider each term one at a time. The first term $B(\eta,\xi)$ is straightforwardly bounded by:

$$|B(\eta,\xi)| \leq \|\eta\|_{e,\mathcal{T}_h}\|\xi\|_{e,\mathcal{T}_h} \leq C\frac{h^{\mu-1}}{p^{s-1}}\|u\|_s\|\xi\|_{e,\mathcal{T}_h} \tag{4.5}$$

We expect that the third term $J(\xi,\eta)$ can be treated as before and should not pose any problems. Indeed, applying the Cauchy-Schwartz inequality, we have:

$$|J(\xi,\eta)| \leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \langle \mathbf{n}\cdot\nabla\xi\rangle^2 \,\mathrm{d}s}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} [\eta]^2 \,\mathrm{d}s}$$

When $\gamma \subset \Gamma_{int}\cup\Gamma_D$ and $\xi \in \mathcal{V}^{hp}(K)$, we have already shown that:

$$\int_\gamma (\mathbf{n}\cdot\nabla\xi)^2 \,\mathrm{d}s \leq C\frac{p_K^2}{h_K}\|\nabla\xi\|_{0,K}^2.$$

Next, we obtain from the approximation property (A.9)

$$\int_\gamma \eta^2 \,\mathrm{d}s = \|\eta\|_{0,\gamma}^2 \leq C\frac{h_K^{2\mu-1}}{p_K^{2s-1}}\|u\|_{s,K}^2.$$

Therefore the term $J(\xi,\eta)$ is bounded by:

$$|J(\xi,\eta)| \leq C\frac{h^{\mu-1}}{p^{s-3/2}}\|u\|_s\|\xi\|_{e,\mathcal{T}_h}.$$

Finally we need to consider the term $J(\eta,\xi)$, which is held responsible for deteriorating the convergence rate of the solution. By the Cauchy-Schwarz inequality, we have:

$$|J(\eta,\xi)| \leq \sqrt{\int_{\Gamma_{int}\cup\Gamma_D} \langle \mathbf{n}\cdot\nabla\eta\rangle^2 \,\mathrm{d}s}\sqrt{\int_{\Gamma_{int}\cup\Gamma_D} [\xi]^2 \,\mathrm{d}s}$$

Once again, the approximation property gives

$$\int_\gamma (\mathbf{n}\cdot\nabla\eta)^2 \,\mathrm{d}s \leq C\frac{h_K^{2\mu-3}}{p_K^{2s-3}}\|u\|_{s,K}^2$$

while from the trace inequality (A.3), we have:

$$\begin{aligned}
\|\xi\|_{0,\gamma}^2 &\leq C\left(\frac{1}{h_K}\|\xi\|_{0,K}^2 + \|\xi\|_{0,K}\|\nabla\xi\|_{0,K}\right) \\
&\leq C\left(\frac{1}{h_K}\|\xi\|_{0,K}^2 + \frac{1}{h_K}\|\xi\|_{0,K}^2 + h_K\|\nabla\xi\|_{0,K}^2\right) \\
&\leq C\left(\frac{1}{h_K}\|\xi\|_{0,K}^2 + h_K\|\nabla\xi\|_{0,K}^2\right) \qquad (4.6)\\
&\leq C\left(\frac{1}{ch_K}\|\xi\|_{e,K}^2 + h_K\|\xi\|_{e,K}^2\right) \\
&\leq \frac{C}{ch_K}\|\xi\|_{e,K}^2.
\end{aligned}$$

It is important to point out here that the norm $\|\xi\|_{0,\gamma}$ is bounded as long as $c > 0$. Then we have:

$$|J(\eta,\xi)| \leq C\frac{h^{\mu-2}}{p^{s-3/2}}\|u\|_s\|\xi\|_{e,\mathcal{T}_h}.$$

In conclusion,

$$\|\xi\|_{e,\mathcal{T}_h} \leq C\left(\frac{h^{\mu-1}}{p^{s-1}} + \frac{h^{\mu-1}}{p^{s-3/2}} + \frac{h^{\mu-2}}{p^{s-3/2}}\right)\|u\|_s \leq C\frac{h^{\mu-2}}{p^{s-3/2}}\|u\|_s$$

which completes the proof. □

Remark 6 *Note that C is inversely proportional to c. Therefore the error is expected to grow as c gets smaller.*

4.2.2. Discussion of the case in which c is zero

The operator, when c is zero, reduces to the pure Laplacian. In this case, the energy norm $\|\cdot\|_{e,\mathcal{T}_h}$ becomes the seminorm $\|\nabla\cdot\|_{0,\mathcal{T}_h}$. Following the same procedure as before, we would have:

$$\|\nabla\xi\|^2_{0,\mathcal{T}_h} = \mathcal{B}_+(\xi,\xi) = \mathcal{B}_+(\eta,\xi) \tag{4.7}$$

where $\eta = u - z_p$, $\xi = u_h - z_p$ and z_p defines an arbitrary interpolant of u on $\mathcal{V}^{hp}$. However, from (4.6), we can see right now that the term $\mathcal{B}_+(\eta,\xi)$ would then be bounded by $\|\xi\|_{0,\mathcal{T}_h}$. In turn, it is impossible to bound $\|\xi\|_{0,\mathcal{T}_h}$ with respect to $\|\nabla\xi\|_{0,\mathcal{T}_h}$. Therefore, the previous methodology to obtain error estimates cannot be applied in the present case.

Suppose that we introduce an elementwise constant function $\bar{\xi}$ to be defined later. Then, we can rewrite (4.7) as:

$$\|\nabla\xi\|^2_{0,\mathcal{T}_h} = \mathcal{B}_+(\eta,\xi) = \mathcal{B}_+(\eta,\xi-\bar{\xi}+\bar{\xi}) = \mathcal{B}_+(\eta,\xi-\bar{\xi}) + \mathcal{B}_+(\eta,\bar{\xi}). \tag{4.8}$$

Suppose now we can construct a new interpolant such that:

$$\mathcal{B}_+(\eta,\bar{\xi}) = 0. \tag{4.9}$$

Then we would have

$$\begin{aligned}\|\nabla\xi\|^2_{0,\mathcal{T}_h} &= \mathcal{B}_+(\eta,\xi) = \mathcal{B}_+(\eta,\xi-\bar{\xi})\\ &= B(\eta,\xi-\bar{\xi}) - J(\eta,\xi-\bar{\xi}) + J(\xi-\bar{\xi},\eta)\\ &= B(\eta,\xi) - J(\eta,\xi-\bar{\xi}) + J(\xi,\eta)\end{aligned} \tag{4.10}$$

We have seen that the terms $B(\eta,\xi)$ and $J(\xi,\eta)$ are easily bounded in terms of $\|\nabla\xi\|_{0,\mathcal{T}_h}$. The other term reads:

$$J(\eta,\xi-\bar{\xi}) = \int_{\Gamma_{int}\cup\Gamma_D} \langle \mathbf{n}\cdot\nabla\eta\rangle\,[\xi-\bar{\xi}]\,\mathrm{d}s.$$

According to Lemma A.5, this integral can be bounded with respect to $\|\nabla\xi\|_{0,\mathcal{T}_h}$ under the condition that $\bar{\xi}$ is chosen as the average of ξ on each element.

This approach has been followed in principle by Rivière, Wheeler and Girault in [20,19] where they construct special interpolants πu which satisfied (4.9) and

$$\|u-\pi u\|_{0,K} \le C\frac{h_K^{\mu}}{p_K^{s-2}}\|u\|_{s,K},$$

$$\|\nabla(u-\pi u)\|_{0,K} \le C\frac{h_K^{\mu-1}}{p_K^{s-2}}\|u\|_{s,K},$$

$$\|\nabla^2(u-\pi u)\|_{0,K} \le C\frac{h_K^{\mu-2}}{p_K^{s-2}}\|u\|_{s,K},$$

where $\mu=\min(p_K+1,s)$, $s\ge 2$, $p_K\ge 2$. Using these interpolants, they were able to derive an *a priori* error estimate of the form:

$$\|\nabla e\|_{0,\mathcal{T}_h} \le C\frac{h^{\mu-1}}{p^{s-4}}\|u\|_s. \tag{4.11}$$

Although the rate of convergence is optimal in h, we show next that the rate of convergence in p is in reality better than $(s-4)$. We improve this result by constructing better approximation properties for the new interpolant and by refining the analysis.

4.2.3. New Interpolants

Lemma 4.1 *Let K be a triangular element of the partition $\mathcal{T}_h$ and u a function in $H^s(K)$, $s\ge 2$. There exists a positive constant C depending on s and ϱ but independent of u, p_K, and h_K, and a polynomial $\pi u\in P_{p_K}(K)$, $p_K\ge 2$, such that*

$$\int_\gamma \mathbf{n}\cdot\nabla(u-\pi u)\,\mathrm{d}s = 0, \qquad \forall\gamma\subset\partial K, \tag{4.12}$$

and

$$\begin{aligned}
\|u-\pi u\|_{0,K} &\le C\frac{h_K^{\mu}}{p_K^{s-3/2}}\|u\|_{s,K},\\
\|\nabla(u-\pi u)\|_{0,K} &\le C\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\|u\|_{s,K},\\
\|\nabla^2(u-\pi u)\|_{0,K} &\le C\frac{h_K^{\mu-2}}{p_K^{s-2}}\|u\|_{s,K},
\end{aligned} \tag{4.13}$$

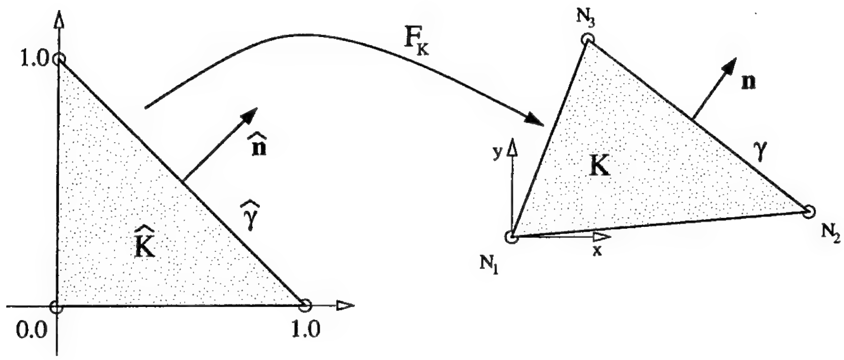

Figure 2. Reference element $\hat{K}$ and mapping F_K from $\hat{K}$ to the element K in the physical domain.

where $\mu = \min(p_K + 1, s)$.

We present the proof of this theorem for triangular elements only. The proof is similar for quadrilaterals.

Proof: Let the triangle $K \in \Omega$ be the image of the master element $\hat{K}$ by the **affine** mapping F_K as shown in Figure 2. The mapping F_K is often rewritten as:

$$F_K(\hat{\mathbf{x}}) = B\hat{\mathbf{x}} + \boldsymbol{b} \tag{4.14}$$

where B represents a two-by-two matrix whose components are independent of $\hat{\mathbf{x}}$ and $\boldsymbol{b}$ is a two-dimensional vector. Here, γ will refer to the edge between node N_2 and N_3, unless stated otherwise, and $\hat{\gamma}$ on $\hat{K}$ will denote its image by F_K^{-1}. We associate with $\hat{\gamma}$ and γ the unit normal vector $\hat{\mathbf{n}}$ and $\mathbf{n}$, respectively.

Given $\eta \in H^2(K)$, namely $\eta = u - z_p$, where z_p is the interpolant of u as defined in Lemma A.7, the objective here is to construct a polynomial function q in $\mathcal{V}^{hp}(K)$ such that:

$$\int_\gamma \mathbf{n} \cdot \nabla \eta \, \mathrm{d}s = \int_\gamma \mathbf{n} \cdot \nabla q \, \mathrm{d}s. \tag{4.15}$$

Indeed we would have:

$$\int_\gamma \mathbf{n} \cdot \nabla(\eta - q) \, \mathrm{d}s = \int_\gamma \mathbf{n} \cdot \nabla(u - z_p - q) \, \mathrm{d}s = \int_\gamma \mathbf{n} \cdot \nabla(u - (z_p + q)) \, \mathrm{d}s = 0,$$

and the new interpolant could be derived as $\pi u = z_p + q$.

Following [19], and assuming $p_K \geq 2$, we introduce the polynomial function $\hat{q}_\gamma$ associated with the edge $\hat{\gamma}$ on $\hat{K}$:

$$\hat{q}_\gamma = C_\gamma (1 - \hat{x} - \hat{y})(\hat{x} + \hat{y}), \qquad \forall\, \hat{\mathbf{x}} = (\hat{x}, \hat{y}) \in \hat{K}. \tag{4.16}$$

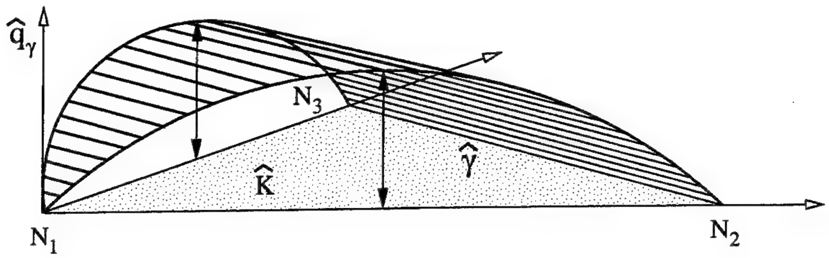

Figure 3. Polynomial function $\hat{q}_\gamma$ on the reference element $\hat{K}$.

where C_γ is a constant to be defined. We observe in Fig. 3 that such a polynomial function satisfies:

$$\int_{\hat{\gamma}_{12}} \hat{\mathbf{n}} \cdot \hat{\nabla}\hat{q}_\gamma \, ds = 0$$

$$\int_{\hat{\gamma}_{23}=\hat{\gamma}} \hat{\mathbf{n}} \cdot \hat{\nabla}\hat{q}_\gamma \, ds = -2C_\gamma$$

$$\int_{\hat{\gamma}_{31}} \hat{\mathbf{n}} \cdot \hat{\nabla}\hat{q}_\gamma \, ds = 0$$

with $\hat{\gamma}_{ij}$ defining an edge on $\hat{K}$ joining the nodes N_i and N_j.

The constant C_γ is found so that (4.15) is satisfied in the physical space. Furthermore, we obtain an upper bound for C_γ (see Rivière [19]) as:

$$\begin{aligned}
|C_\gamma| &\leq C\|B\|^2\|B^{-1}\|^2\|\hat{\nabla}\hat{\eta}\|_{0,\hat{\gamma}} \\
&\leq C\left(\frac{h_K}{\hat{\rho}}\right)^2\left(\frac{\hat{h}}{\rho_K}\right)^2 h_K^{1/2}\,\|\nabla\eta\|_{0,\gamma} \\
&\leq C\sigma^2\left(\frac{\hat{h}}{\hat{\rho}}\right)^2 h_K^{1/2}\,\|\nabla\eta\|_{0,\gamma} \\
&\leq C h_K^{1/2}\,\|\nabla\eta\|_{0,\gamma} \\
&\leq C\left\{\|\nabla\eta\|_{0,K}^2 + h_K\|\nabla\eta\|_{0,K}\|\nabla^2\eta\|_{0,K}\right\}^{1/2}
\end{aligned}$$

where we make use of the Trace Inequality (A.3). We also observe that:

$$\begin{aligned}
\|q_\gamma\|_{0,K} &\leq C|\det B|^{1/2}\|\hat{q_\gamma}\|_{0,\hat{K}} \\
&\leq C h_K|C_\gamma||\hat{K}| \\
&\leq C h_K|C_\gamma|.
\end{aligned}$$

Likewise, we have:

$$\|\nabla q_\gamma\|_{0,K} \le C|C_\gamma|$$
$$\|\nabla^2 q_\gamma\|_{0,K} \le C h_K^{-1}|C_\gamma|$$

So far, we have carried out the analysis for the edge γ between node N_2 and N_3. We point out that the same results are obtained for the other two edges. We then associate with each edge γ_{12}, γ_{23}, γ_{31}, a polynomial q_{12}, q_{23}, q_{31} respectively such that

$$\hat{q}_{12} = C_{12}\,\hat{y}(1-\hat{y})$$
$$\hat{q}_{23} = C_{23}\,(1-\hat{x}-\hat{y})(\hat{x}+\hat{y})$$
$$\hat{q}_{31} = C_{31}\,\hat{x}(1-\hat{x})$$

Adding these polynomial functions together, we construct on the element K a new function $q \in P_2(K)$

$$q(\mathbf{x}) = q_{12}(\mathbf{x}) + q_{23}(\mathbf{x}) + q_{31}(\mathbf{x}), \qquad \forall \mathbf{x} \in K,$$

which satisfies

$$\int_{\gamma_{12}} \mathbf{n}\cdot\nabla q\,\mathrm{d}s = \int_{\gamma_{12}} \mathbf{n}\cdot\nabla(q_{12}+q_{23}+q_{31})\,\mathrm{d}s = \int_{\gamma_{12}} \mathbf{n}\cdot\nabla q_{12}\,\mathrm{d}s = \int_{\gamma_{12}} \mathbf{n}\cdot\nabla\eta\,\mathrm{d}s,$$
$$\int_{\gamma_{23}} \mathbf{n}\cdot\nabla q\,\mathrm{d}s = \int_{\gamma_{23}} \mathbf{n}\cdot\nabla(q_{12}+q_{23}+q_{31})\,\mathrm{d}s = \int_{\gamma_{23}} \mathbf{n}\cdot\nabla q_{23}\,\mathrm{d}s = \int_{\gamma_{23}} \mathbf{n}\cdot\nabla\eta\,\mathrm{d}s,$$
$$\int_{\gamma_{31}} \mathbf{n}\cdot\nabla q\,\mathrm{d}s = \int_{\gamma_{31}} \mathbf{n}\cdot\nabla(q_{12}+q_{23}+q_{31})\,\mathrm{d}s = \int_{\gamma_{31}} \mathbf{n}\cdot\nabla q_{31}\,\mathrm{d}s = \int_{\gamma_{31}} \mathbf{n}\cdot\nabla\eta\,\mathrm{d}s.$$

In other words, there exists a function $\pi u \in P_P(K)$, $\pi u = z_p + q$ such that

$$\int_\gamma \mathbf{n}\cdot\nabla(u-\pi u)\,\mathrm{d}s = 0, \qquad \forall\gamma \subset \partial K.$$

Now, by the triangle inequality,

$$
\begin{aligned}
\|u-\pi u\|_{0,K} &\le \|u-z_p\|_{0,K}+\|q\|_{0,K}\\
&\le \|\eta\|_{0,K}+\|q_{12}\|_{0,K}+\|q_{23}\|_{0,K}+\|q_{31}\|_{0,K}\\
&\le \|\eta\|_{0,K}+Ch_K\left\{\|\nabla\eta\|_{0,K}^2+h_K\|\nabla\eta\|_{0,K}\|\nabla^2\eta\|_{0,K}\right\}^{1/2}\\
&\le C\left\{\frac{h_K^{\mu}}{p_K^{s}}+h_K\left(\frac{h_K^{2\mu-2}}{p_K^{2s-2}}+h_K\frac{h_K^{\mu-1}}{p_K^{s-1}}\frac{h_K^{\mu-2}}{p_K^{s-2}}\right)^{1/2}\right\}\|u\|_{s,K}\\
&\le C\left\{\frac{h_K^{\mu}}{p_K^{s}}+h_K\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\right\}\|u\|_{s,K}\\
&\le C\frac{h_K^{\mu}}{p_K^{s-3/2}}\|u\|_{s,K}.
\end{aligned}
$$

In the same manner, we find:

$$
\begin{aligned}
\|\nabla(u-\pi u)\|_{0,K} &\le \|\nabla(u-z_p)\|_{0,K}+\|\nabla q\|_{0,K}\\
&\le \|\nabla\eta\|_{0,K}+C\left\{\|\nabla\eta\|_{0,K}^2+h_K\|\nabla\eta\|_{0,K}\|\nabla^2\eta\|_{0,K}\right\}^{1/2}\\
&\le C\left\{\frac{h_K^{\mu-1}}{p_K^{s-1}}+\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\right\}\|u\|_{s,K}\\
&\le C\frac{h^{\mu-1}}{p^{s-3/2}}\|u\|_{s,K},
\end{aligned}
$$

and

$$
\begin{aligned}
\|\nabla^2(u-\pi u)\|_{0,K} &\le \|\nabla^2(u-z_p)\|_{0,K}+\|\nabla^2 q\|_{0,K}\\
&\le \|\nabla^2\eta\|_{0,K}+Ch_K^{-1}\left\{\|\nabla\eta\|_{0,K}^2+h_K\|\nabla\eta\|_{0,K}\|\nabla^2\eta\|_{0,K}\right\}^{1/2}\\
&\le C\left\{\frac{h_K^{\mu-2}}{p_K^{s-2}}+h_K^{-1}\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\right\}\|u\|_{s,K}\\
&\le C\frac{h^{\mu-2}}{p^{s-2}}\|u\|_{s,K}.
\end{aligned}
$$

We observe that the first two estimates are governed by the rate of convergence of $\|q\|_{0,K}$ and $\|\nabla q\|_{0,K}$ respectively, while the last estimate is governed by the rate of convergence of $\|\nabla^2\eta\|_{0,K}$. □

4.2.4. *A priori* error estimate when c is zero

Theorem 4.3 *Let $u \in H^1(\Omega) \cap H^s(\mathcal{T}_h)$, $s \geq 2$ be a solution of (3.23) and u_h be the discrete discontinuous solution of (3.24) with $c = 0$ and $p \geq 2$. Then, the numerical error $e = u - u_h$ satisfies:*

$$||\nabla e||_{0,\mathcal{T}_h} \leq C \frac{h^{\mu-1}}{p^{s-5/2}} ||u||_s \tag{4.17}$$

where $\mu = \min(p+1, s)$.

Proof: Let πu be the interpolant of u in $\mathcal{V}^{hp}$, defined on each element K of $\mathcal{T}_h$ as in Lemma 4.1. We also introduce $\eta = u - \pi u$ and $\xi = u_h - \pi u$. Using the triangle inequality, we have:

$$||\nabla e||_{0,\mathcal{T}_h} = ||\nabla(\eta - \xi)||_{0,\mathcal{T}_h} \leq ||\nabla \eta||_{0,\mathcal{T}_h} + ||\nabla \xi||_{0,\mathcal{T}_h}$$

and from (4.7) and (4.8), we recall that:

$$||\nabla \xi||^2_{0,\mathcal{T}_h} = \mathcal{B}_+(\eta, \xi) = \mathcal{B}_+(\eta, \xi - \bar{\xi}) + \mathcal{B}_+(\eta, \bar{\xi}).$$

Here $\bar{\xi}$ is chosen as the average of ξ over each K, i.e.

$$\bar{\xi} = \frac{1}{|K|} \int_K \xi \, \mathrm{d}x, \qquad K \in \mathcal{T}_h.$$

We note here that the authors in [20,19] chose $\bar{\xi}$ as the average of ξ over each edge and their proof is thus slightly different from ours.

This particular choice of the interpolant πu and piecewise constant function $\bar{\xi}$ does indeed yield:

$$\begin{aligned}
\mathcal{B}_+(\eta, \bar{\xi}) &= B(\eta, \bar{\xi}) - J(\eta, \bar{\xi}) + J(\bar{\xi}, \eta) = -J(\eta, \bar{\xi}) \\
&= -\int_{\Gamma_{int} \cup \Gamma_D} \langle \mathbf{n} \cdot \nabla \eta \rangle \, [\bar{\xi}] \, \mathrm{d}s \\
&= -[\bar{\xi}] \int_{\Gamma_{int} \cup \Gamma_D} \langle \mathbf{n} \cdot \nabla \eta \rangle \, \mathrm{d}s \\
&= 0
\end{aligned}$$

since the last integral is zero according to the property (4.12) of the interpolant πu. Therefore

$$||\nabla \xi||^2_{0,\mathcal{T}_h} = \mathcal{B}_+(\eta, \xi - \bar{\xi}) \tag{4.18}$$

We now show how $\mathcal{B}_+(\eta, \xi - \bar{\xi})$ can be bounded with respect to $||\nabla \xi||_{0,\mathcal{T}_h}$. We naturally have from (4.10)

$$|\mathcal{B}_+(\eta, \xi - \bar{\xi})| \leq |B(\eta, \xi)| + |J(\eta, \xi - \bar{\xi})| + |J(\xi, \eta)|$$

The first term gives, using the approximation properties of Lemma 4.1 and the discrete Schwarz inequality:

$$
\begin{aligned}
|B(\eta,\xi)| &\leq \sum_{K\in\mathcal{T}_h}\int_K |\nabla\eta\cdot\nabla\xi|\,\mathrm{d}x \leq \sum_{K\in\mathcal{T}_h}\|\nabla\eta\|_{0,K}\|\nabla\xi\|_{0,K} \\
&\leq \sum_{K\in\mathcal{T}_h} C\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\|u\|_{s,K}\|\nabla\xi\|_{0,K} \\
&\leq C\frac{h^{\mu-1}}{p^{s-3/2}}\|u\|_s\|\nabla\xi\|_{0,\mathcal{T}_h}
\end{aligned}
$$

The third term is treated as usual. We have

$$
\begin{aligned}
|J(\xi,\eta)| &\leq \int_{\Gamma_{int}\cup\Gamma_D}|\langle\mathbf{n}\cdot\nabla\xi\rangle\,[\eta]|\,\mathrm{d}s \leq \sum_{\gamma}\|\langle\mathbf{n}\cdot\nabla\xi\rangle\|_{0,\gamma}\|[\eta]\|_{0,\gamma} \\
&\leq C\sum_{K\in\mathcal{T}_h}\sum_{\gamma\in\partial K\setminus\Gamma_N}\|\mathbf{n}\cdot\nabla\xi\|_{0,\gamma}\|\eta\|_{0,\gamma} \\
&\leq C\sum_{K\in\mathcal{T}_h}\sum_{\gamma\in\partial K\setminus\Gamma_N}\|\nabla\xi\|_{0,\gamma}\|\eta\|_{0,\gamma}
\end{aligned}
$$

From the trace inequality (A.3) and the inverse property (A.7), we show that:

$$
\begin{aligned}
\|\nabla\xi\|_{0,\gamma} &\leq C\left\{\frac{1}{h_K}\|\nabla\xi\|_{0,K}^2+\|\nabla\xi\|_{0,K}\|\nabla^2\xi\|_{0,K}\right\}^{1/2} \\
&\leq C\left\{\frac{1}{h_K}\|\nabla\xi\|_{0,K}^2+\|\nabla\xi\|_{0,K}C_0\frac{p_K^2}{h_K}\|\nabla\xi\|_{0,K}\right\}^{1/2} \\
&\leq C\frac{p_K}{h_K^{1/2}}\|\nabla\xi\|_{0,K}
\end{aligned}
$$

and, from the approximation properties of Lemma 4.1:

$$
\begin{aligned}
\|\eta\|_{0,\gamma} &\leq C\left\{\frac{1}{h_K}\|\eta\|_{0,K}^2+\|\eta\|_{0,K}\|\nabla\eta\|_{0,K}\right\}^{1/2} \\
&\leq C\left\{\frac{1}{h_K}\frac{h_K^{2\mu}}{p_K^{2s-3}}\|u\|_{s,K}^2+\frac{h_K^{\mu}}{p_K^{s-3/2}}\frac{h_K^{\mu-1}}{p_K^{s-3/2}}\|u\|_{s,K}^2\right\}^{1/2} \\
&\leq C\left\{\frac{h_K^{2\mu-1}}{p_K^{2s-3}}+\frac{h_K^{2\mu-1}}{p_K^{2s-3}}\right\}^{1/2}\|u\|_{s,K} \\
&\leq C\frac{h_K^{\mu-1/2}}{p_K^{s-3/2}}\|u\|_{s,K}
\end{aligned}
$$

In conclusion, we find that:

$$
\begin{aligned}
|J(\xi,\eta)| &\le C \sum_{K\in\mathcal{T}_h} \frac{p_K}{h_K^{1/2}} \|\nabla\xi\|_{0,K} \frac{h_K^{\mu-1/2}}{p_K^{s-3/2}} \|u\|_{s,K} \\
&\le C \sum_{K\in\mathcal{T}_h} \frac{h_K^{\mu-1}}{p_K^{s-5/2}} \|\nabla\xi\|_{0,K} \|u\|_{s,K} \\
&\le C \frac{h^{\mu-1}}{p^{s-5/2}} \|u\|_s \|\nabla\xi\|_{0,\mathcal{T}_h}.
\end{aligned}
$$

In the same manner as before, we obtain for the term $J(\eta,\xi-\bar{\xi})$:

$$
|J(\eta,\xi-\bar{\xi})| \le C \sum_{K\in\mathcal{T}_h} \sum_{\gamma\in\partial K\setminus\Gamma_N} \|\nabla\eta\|_{0,\gamma} \|\xi-\bar{\xi}\|_{0,\gamma}
$$

In this case, we have using also the approximation properties of the interpolant:

$$
\begin{aligned}
\|\nabla\eta\|_{0,\gamma} &\le C \left\{ \frac{1}{h_K} \|\nabla\eta\|_{0,K}^2 + \|\nabla\eta\|_{0,K} \|\nabla^2\eta\|_{0,K} \right\}^{1/2} \\
&\le C \left\{ \frac{1}{h_K} \frac{h_K^{2\mu-2}}{p_K^{2s-3}} \|u\|_{s,K}^2 + \frac{h_K^{\mu-1}}{p_K^{s-3/2}} \frac{h_K^{\mu-2}}{p_K^{s-2}} \|u\|_{s,K}^2 \right\}^{1/2} \\
&\le C \left\{ \frac{h_K^{2\mu-3}}{p_K^{2s-3}} + \frac{h_K^{2\mu-2}}{p_K^{2s-7/2}} \right\}^{1/2} \|u\|_{s,K} \\
&\le C \frac{h_K^{\mu-3/2}}{p_K^{s-7/4}} \|u\|_{s,K}
\end{aligned}
$$

However, for the other term, we have, using Lemma A.5:

$$
\begin{aligned}
\|\xi-\bar{\xi}\|_{0,\gamma} &\le C \left\{ \frac{1}{h_K} \|\xi-\bar{\xi}\|_{0,K}^2 + \|\xi-\bar{\xi}\|_{0,K} \|\nabla(\xi-\bar{\xi})\|_{0,K} \right\}^{1/2} \\
&\le C \left\{ \frac{1}{h_K} \|\xi-\bar{\xi}\|_{0,K}^2 + \|\xi-\bar{\xi}\|_{0,K} \|\nabla\xi\|_{0,K} \right\}^{1/2} \\
&\le C \left\{ \frac{1}{h_K} h_K^2 \|\nabla\xi\|_{0,K}^2 + h_K \|\nabla\xi\|_{0,K} \|\nabla\xi\|_{0,K} \right\}^{1/2} \\
&\le C h_K^{1/2} \|\nabla\xi\|_{0,K}
\end{aligned}
$$

It follows that:

$$
\begin{aligned}
|J(\eta,\xi-\bar{\xi})| &\leq C \sum_{K\in\mathcal{T}_h} \frac{h_K^{\mu-3/2}}{p_K^{s-7/4}} \|u\|_{s,K}\, h_K^{1/2} \|\nabla\xi\|_{0,K} \\
&\leq C \sum_{K\in\mathcal{T}_h} \frac{h_K^{\mu-1}}{p_K^{s-7/4}} \|u\|_{s,K} \|\nabla\xi\|_{0,K} \\
&\leq C \frac{h^{\mu-1}}{p^{s-7/4}} \|u\|_s \|\nabla\xi\|_{0,\mathcal{T}_h}
\end{aligned}
$$

Combining the previous results, we finally get:

$$
\|\nabla\xi\|_{0,\mathcal{T}_h} \leq C\left(\frac{h^{\mu-1}}{p^{s-1/2}} + \frac{h^{\mu-1}}{p^{s-7/4}} + \frac{h^{\mu-1}}{p^{s-5/2}}\right)\|u\|_s \leq C\frac{h^{\mu-1}}{p^{s-5/2}}\|u\|_s
$$

and this completes the proof since $\|\nabla\eta\|_{0,\mathcal{T}_h}$ converges with a greater rate of convergence than $\|\nabla\xi\|_{0,\mathcal{T}_h}$. □

4.2.5. Alternative estimate when c is nonzero

We now use the previous results to review the error estimate when c is nonzero. The new estimate is given in the following theorem:

Theorem 4.4 *Let $u \in H^1(\Omega)\cap H^s(\mathcal{T}_h)$, $s \geq 2$, be a solution of (3.23) with $c > 0$ and u_h be the discrete discontinuous solution of (3.24). Then, the numerical error $e = u - u_h$ satisfies:*

$$
\|e\|_{e,\mathcal{T}_h} \leq C\frac{h^{\mu-1}}{p^{s-5/2}}\|u\|_s \tag{4.19}
$$

where $\mu = \min(p+1, s)$ and $p \geq 2$.

Proof: In this case, we have:

$$
\begin{aligned}
\|\xi\|^2_{e,\mathcal{T}_h} &= \mathcal{B}_+(\eta,\xi) \\
&= B(\eta,\xi) - J(\eta,\xi) + J(\xi,\eta) \\
&= B(\eta,\xi) - J(\eta,\xi-\bar{\xi}) - J(\eta,\hat{\xi}) + J(\xi,\eta) \\
&= B(\eta,\xi) - J(\eta,\xi-\bar{\xi}) + J(\xi,\eta) \\
&\leq |B(\eta,\xi)| + |J(\eta,\xi-\bar{\xi})| + |J(\xi,\eta)|
\end{aligned}
$$

if the interpolant is chosen as in Lemma 4.1.

Moreover, results from the previous theorem provide us with:

$$
\begin{aligned}
|B(\eta,\xi)| &\le \sum_{K\in\mathcal{P}_h}\int_K |\nabla\eta\cdot\nabla\xi + c\eta\xi|\,\mathrm{d}x \le C\frac{h^{\mu-1}}{p^{s-3/2}}\|u\|_s\|\xi\|_{e,\mathcal{P}_h},\\
|J(\xi,\eta)| &\le C\frac{h^{\mu-1}}{p^{s-5/2}}\|u\|_s\|\nabla\xi\|_{0,\mathcal{P}_h} \le C\frac{h^{\mu-1}}{p^{s-5/2}}\|u\|_s\|\xi\|_{e,\mathcal{P}_h},\\
|J(\eta,\xi)| &\le C\frac{h^{\mu-1}}{p^{s-7/4}}\|u\|_s\|\nabla\xi\|_{0,\mathcal{P}_h} \le C\frac{h^{\mu-1}}{p^{s-7/4}}\|u\|_s\|\xi\|_{e,\mathcal{P}_h},
\end{aligned}
$$

so that

$$
\|\xi\|_{e,\mathcal{P}_h} \le C\frac{h^{\mu-1}}{p^{s-5/2}}\|u\|_s,
$$

and this completes the proof. □

This time, the rate of convergence is optimal with respect to h but the rate of convergence in p is worse than in the previous estimate. This makes us believe that the error estimates for the DG method can still be improved with respect to p. Maybe better interpolants are yet to be found.

5. Concluding Remarks

5.1. Remarks on the Discontinuous Formulations

We have studied here four different formulations of the so-called Discontinuous Galerkin Method (DGM). These formulations simply vary by one sign (plus or minus) and by the addition of a penalty term (or not). However, they greatly differ in nature from a mathematical point of view. We now review each formulation one by one and recount our findings in the case of linear diffusion problems.

Global Element Method. Little can be proved for this method. We were able to derive the continuity of the associated bilinear form, but failed to even obtain *a priori* error estimates. This is because the bilinear form is not guaranteed to be semi-positive definite.

Symmetric Interior Penalty Galerkin Method. The SIPG Method is similar to the GEM except for the addition of the penalty term. However, it allows us to prove non only continuity of the bilinear form, but also coercivity in the discrete discontinuous space (for sufficiently large values of the penalty parameter), and thus *a priori* error estimates optimal with respect to h $(\mu-1)$ and slighty suboptimal with respect to p $(s-3/2)$. One major drawback of this method is that its behavior depends on the selection of the penalty parameter. If not chosen carefully, the method can fail.

Non-Symmetric Interior Penalty Galerkin Method. The limitation of the SIPG method is remedied by changing one minus sign by a plus sign. Indeed, although the NIPG formulation results in a non-symmetric system of equations, all the properties and error estimates are shown to be independent of the choice of the penalty parameter. We also find the same rates of convergence with respect to h and p as SIPG.

Discontinuous Galerkin Method. DGM is deduced from the NIPG method by setting the penalty parameter to zero. We then observe that the rate of convergence with respect to h or p deteriorates. Also, in the case of the pure Laplacian operator, when c is set to zero in the Poisson problem, we obtain *a priori* error estimates only by defining some new interpolants whose fluxes are weakly equal to the fluxes of the exact solution over each edge of the elements. Although the rate of convergence in h remains optimal, the one in p is then estimated to be $s - 5/2$. We believe that it might be possible to improve this rate of convergence by considering other types of interpolants. At this point, detailed numerical experiments would be helpful to understand how the penalty term affects the quality of the approximations.

5.2. Future Challenges

The great challenges for DGMs are to 1) prove uniqueness of the solutions of the continuous formulations, 2) perform more numerical experiments to understand the role played by the penalty terms, 3) still improve the *a priori* error estimates for the Discontinuous Galerkin Method of Baumann and Oden, 4) derive rigorous *a posteriori* error estimates for the various formulations.

A. Lemmas

A.1. Discrete Schwarz Inequality

Lemma A.1 *Let $\{a_i\}$ and $\{b_i\}$ define two sequences of N real numbers. Then*

$$\sum_{i=1}^{N} a_i b_i \leq \left(\sum_{i=1}^{N} a_i^2\right)^{1/2} \left(\sum_{i=1}^{N} b_i^2\right)^{1/2} \tag{A.1}$$

Proof: We shall show the discrete Schwarz inequality for $N = 2$ first. We have:

$$\begin{aligned}
(a_1 b_1 + a_2 b_2)^2 &= a_1^2 b_1^2 + a_2^2 b_2^2 + 2 a_1 b_1 a_2 b_2 \\
&= (a_1^2 + a_2^2)(b_1^2 + b_2^2) - a_1^2 b_2^2 - a_2^2 b_1^2 + 2 a_1 b_1 a_2 b_2 \\
&= (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_2 - a_2 b_1)^2 \\
&\leq (a_1^2 + a_2^2)(b_1^2 + b_2^2)
\end{aligned}$$

so that:

$$a_1b_1 + a_2b_2 \leq \sqrt{a_1^2 + a_2^2}\sqrt{b_1^2 + b_2^2}.$$

The result is easily extended to $N > 2$ by recursivity. □

A.2. Multiplicative Trace Inequalities

Lemma A.2 *Let Ω define a star-shaped domain with boundary $\partial\Omega$ as shown in Fig. 4. Then, for all $v \in H^1(\Omega)$*

$$\|v\|_{0,\partial\Omega}^2 \leq \frac{2}{\inf\limits_{\mathbf{x}\in\partial\Omega} |\mathbf{x}|}\left(\|v\|_{0,\Omega}^2 + \sup_{\mathbf{x}\in\Omega} |\mathbf{x}|\,\|v\|_{0,\Omega}\|\nabla v\|_{0,\Omega}\right). \tag{A.2}$$

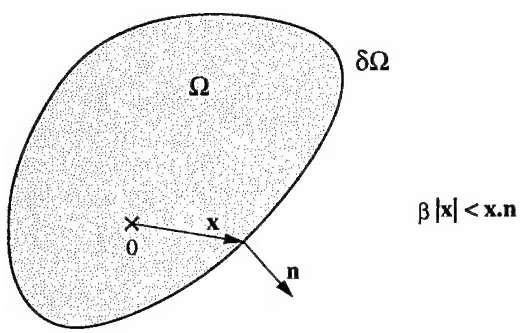

Figure 4. Star-shaped domain.

Proof: Let $O \in \Omega$ be the origin and let $\mathbf{n}$ denote the unit normal outward vector on $\partial\Omega$. From the definition of a star-shaped domain, there exists a positive constant β such that

$$\beta|\mathbf{x}| \leq \mathbf{x}\cdot\mathbf{n}.$$

Applying Green's Theorem for the vector field $u^2\mathbf{x}$, we have:

$$\int_{\partial\Omega} u^2\mathbf{x}\cdot\mathbf{n}\,ds = \int_\Omega \nabla\cdot(u^2\mathbf{x})\,dx.$$

By the property of star-shaped domains, the first integral is shown to be bounded below:

$$\int_{\partial\Omega} u^2\mathbf{x}\cdot\mathbf{n}\,ds \geq \beta \inf_{\mathbf{x}\in\partial\Omega}|\mathbf{x}| \int_{\partial\Omega} u^2\,ds \geq \beta \inf_{\mathbf{x}\in\partial\Omega}|\mathbf{x}|\|u\|_{0,\partial\Omega}^2.$$

On the other hand, the second integral is bounded above:

$$
\begin{aligned}
\int_\Omega \nabla \cdot (u^2 \mathbf{x})\,\mathrm{d}x &= \int_\Omega u^2 \nabla \cdot \mathbf{x} + \mathbf{x} \cdot \nabla u^2 \,\mathrm{d}x \\
&= \int_\Omega 2u^2 \,\mathrm{d}x + \int_\Omega 2u\mathbf{x} \cdot \nabla u \,\mathrm{d}x \\
&\leq 2\|u\|_{0,\Omega}^2 + \int_\Omega |u\mathbf{x} \cdot \nabla u| \,\mathrm{d}x \\
&\leq 2\|u\|_{0,\Omega}^2 + 2 \sup_{\mathbf{x}\in\Omega} |\mathbf{x}| \int_\Omega |u||\nabla u| \,\mathrm{d}x \\
&\leq 2\|u\|_{0,\Omega}^2 + 2 \sup_{\mathbf{x}\in\Omega} |\mathbf{x}| \|u\|_{0,\Omega} \|\nabla u\|_{0,\Omega}
\end{aligned}
$$

Using both bounds, we arrive at:

$$
\|u\|_{0,\partial\Omega}^2 \leq \frac{2}{\inf\limits_{\mathbf{x}\in\partial\Omega} |\mathbf{x}|} \left(\|u\|_{0,\Omega}^2 + \sup_{\mathbf{x}\in\Omega} |\mathbf{x}| \|u\|_{0,\Omega} \|\nabla u\|_{0,\Omega} \right)
$$

which completes the proof. □

Lemma A.3 *Let K be a triangle or a quadrilateral such that $h_K \leq \varrho\rho_K$ (shape regular). Then, for all $v \in H^1(K)$,*

$$
\|v\|_{0,\partial K}^2 \leq C \left(\frac{1}{h_K} \|v\|_{0,K}^2 + \|v\|_{0,K} \|\nabla v\|_{0,K} \right). \tag{A.3}
$$

where C is a positive constant.

Proof: Let the origin O be the center of the inscribed circle in K with radius $\rho_K/2$. We therefore have:

$$
\begin{aligned}
&\sup_{\mathbf{x}\in K} |\mathbf{x}| \leq h_K \\
&\inf_{\mathbf{x}\in\partial K} |\mathbf{x}| \geq \rho_K \geq h_K/\varrho
\end{aligned}
$$

so that from (A.2)

$$
\begin{aligned}
\|u\|_{0,\partial K}^2 &\leq \frac{2\varrho}{h_K} \left(\|u\|_{0,K}^2 + h_K \|u\|_{0,K} \|\nabla u\|_{0,K} \right) \\
&\leq 2\varrho \left(\frac{1}{h_K} \|u\|_{0,K}^2 + \|u\|_{0,K} \|\nabla u\|_{0,K} \right)
\end{aligned}
$$

The proof is complete when choosing $C = 2\varrho$. □

A.3. Poincaré–Friedrich's Inequalities

Lemma A.4 *Let Ω be an open, bounded, connected domain of $\mathbb{R}^2$ with Lipschitz boundary $\partial\Omega$. Let $v \in H^1\Omega$ such that*

$$\int_\Omega v \, \mathrm{d}x = 0. \tag{A.4}$$

Then

$$\|v\|_{0,\Omega} \le C\|\nabla v\|_{0,\Omega} \tag{A.5}$$

where $C = C(\Omega)$ is a positive constant.

Proof: See Schwab [21, p.350] and Brenner and Scott [8, p.102]. □

Lemma A.5 *Let $z \in P_{p_K}(K)$ and $\bar{z}$ be the average of z on K, $\bar{z} = (\int_K z \, \mathrm{d}x)/|K|$. Then*

$$\|z - \bar{z}\|_{0,K} \le Ch_K\|\nabla z\|_{0,K} \tag{A.6}$$

where C is a positive constant independent of K and z.

Proof: Let $z \in P_{p_K}(K)$ and $v = z - \bar{z}$. Then

$$\int_\Omega v \, \mathrm{d}x = \int_\Omega z - \bar{z} \, \mathrm{d}x = \int_\Omega z \, \mathrm{d}x - \int_\Omega \bar{z} \, \mathrm{d}x = |K|\bar{z} - \bar{z}|K| = 0.$$

By a scaling argument and Lemma A.4,

$$\|v\|_{0,K} \le Ch_K\|\hat{v}\|_{0,\hat{K}} \le C(\hat{K})h_K\|\hat{\nabla}\hat{v}\|_{0,\hat{K}} \le Ch_K\|\nabla v\|_{0,K}$$

Substituting $z - \bar{z}$ for v, it follows that $\|z - \bar{z}\|_{0,K} \le Ch_K\|\nabla(z - \bar{z})\|_{0,K}$, in other words, since $\bar{z}$ is constant, $\|z - \bar{z}\|_{0,K} \le Ch_K\|\nabla z\|_{0,K}$. □

A.4. Inverse Property

Lemma A.6 *Let $z \in P_{p_K}(K)$. Then*

$$\|\nabla z\|_{0,K} \le C\frac{p_K^2}{h_K}\|z\|_{0,K} \tag{A.7}$$

Proof: See Schwab [21, p.208]. □

A.5. Interpolation Error Estimates

Lemma A.7 *Let K be a triangle or parallelogram element of the partition $\mathcal{P}_h$ and u a function in $H^s(K)$. There exists a positive constant C depending on s and ϱ but independent of u, p_K, and h_K, and a sequence $z_p \in P_{p_K}(K)$, $p_K = 1, 2, \ldots$, such that for any q, $0 \leq q \leq s$*

$$\|u - z_p\|_{q,K} \leq C \frac{h_K^{\mu-q}}{p_K^{s-q}} \|u\|_{s,K}, \qquad s \geq 0 \tag{A.8}$$

$$\|u - z_p\|_{0,\gamma} \leq C \frac{h_K^{\mu-1/2}}{p_K^{s-1/2}} \|u\|_{s,K}, \qquad s > \frac{1}{2} \tag{A.9}$$

where $\mu = \min(p_K + 1, s)$, $h_K = \text{diam}\,(K)$ *and* $\gamma \subset \partial K$.

Proof: See Babuška and Suri [3,4]. □

REFERENCES

1. D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19:742–, 1978.
2. I. Babuška, C. E. Baumann, and J. T. Oden. A discontinuous hp finite element method for diffusion problems: 1-D analysis. *Computers and Mathematics with Applications*, 37:103–122, 1999.
3. I. Babuška and M. Suri. The h-p version of the finite element method with quasiuniform meshes. *Mathematical Modeling and Numerical Analysis*, 21(2):199–238, 1987.
4. I. Babuška and M. Suri. The optimal convergence rate of the p-version of the finite element method. *SIAM J. Numer. Anal.*, 24(4), 1987.
5. G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Mathematics of Computations*, 31(137):45–59, 1977.
6. G. A. Baker, W. N. Jureidini, and O. A. Karakashian. Piecewise solenoidal vector fields and the Stokes problem. *SIAM J. Numer. Anal.*, 27(6):1466–1485, 1990.
7. C. E. Baumann. *An h-p adaptive discontinuous finite element method for computational fluid dynamics*. PhD thesis, The University of Texas at Austin, 1997.
8. S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New-York, 1994.
9. Z. Chen. On the relationship of various discontinuous finite element methods for second-order elliptic equations. SMU Math Report 2000-02, Southern Methodist University, 2000.
10. B. Cockburn, G. E. Karniadakis, and C. W. Shu, editors. *Discontinuous Galerkin Methods - Theory, Computation and Applications*, volume 11 of *Lectures Notes in Computational Science and Engineering*. Springer, Berlin, 2000.
11. L. M. Delves and C. A. Hall. An implicit matching principle for global element calculations. *J. Inst. Math. Appl.*, 23(2):223–234, 1979.

12. L. M. Delves and C. Phillips. A fast implementation of the global element method. *J. Inst. Math. Appl.*, 25(2):177–197, 1980.
13. J. A. Hendry and L. M. Delves. The global element method applied to a harmonic mixed boundary value problem. *J. Comput. Phys.*, 33:33–, 1979.
14. J. A. Hendry, L. M. Delves, and C. Phillips. Numerical experience with the global element method. In *Mathematics of finite elements and applications, III*, pages 341–348. Academic Press, London-New York, 1979. (Proc. Third MAFELAP Conf., Brunel Univ., Uxbridge, 1978).
15. P. Houston, C. Schwab, and E. Süli. Discontinuous *hp*-finite element methods for advection-diffusion problems. Technical Report no. 00/15, Oxford University Computing Laboratory, 2000.
16. O. A. Karakashian and W. N. Jureidini. A nonconforming finite element method for the stationary Navier-Stokes equations. *SIAM J. Numer. Anal.*, 35(1):93–120, 1998.
17. J. T. Oden, I. Babuška, and C. E. Baumann. A discontinuous *hp* finite element method for diffusion problems. TICAM Report 97-21, The University of Texas at Austin, 1997.
18. J. T. Oden, I. Babuška, and C. E. Baumann. A discontinuous *hp* finite element method for diffusion problems. *J. Comput. Phys.*, 146:491–519, 1998.
19. B. Rivière. *Discontinuous Galerkin Methods for Solving the Miscible Displacement Problem in Porous Media*. PhD thesis, The University of Texas at Austin, May 2000.
20. B. Rivière, M. F. Wheeler, and V. Girault. Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I. *Computational Geosciences*, 3-4:337–360, 1999.
21. C. Schwab. *p- and hp- Finite Element Methods - Theory and Applications in Solid and Fluid Mechanics*. Oxford University Press, New York, 1998.
22. E. Süli, C. Schwab, and P. Houston. *hp*-DGFEM for partial differential equations with nonnegative characteristic form. In B. Cockburn, G. E. Karniadakis, and C. W. Shu, editors, *Discontinuous Galerkin Methods*, volume 11 of *Lectures Notes in Computational Science and Engineering*, pages 221–230. Springer, Berlin, 2000.
23. M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15:152–, 1978.